THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

Section B

VOL. 64, PART 9 1 September 1951 No. 381 B

The Mechanical Properties of Metals

BY N. F. MOTT
H. H. Wills Physical Laboratory, University of Bristol

35th Guthrie Lecture, delivered 16th March 1951; MS. received 25th April 1951

ABSTRACT. A description is given of edge and screw dislocations in a close-packed cubic lattice, which leads up to an account of the theory of Frank and Read of the origin of slip bands. Consideration is then given to a crystalline grain containing Frank–Read sources; it is shown that one has to expect cross-slip and the formation of deformation bands. It is shown moreover that the formation of vacancies by moving dislocations plays an essential role in these processes. Some speculations on the stability of deformation bands are given. It is suggested that the movement of vacancies (self-diffusion) plays an essential part in polygonization, recovery and steady-state creep, enabling dislocations in deformation bands to move out of their slip planes and so relieve stress. It is probable moreover that in these bands the stresses are several hundred times greater than the applied stress, and that this enables diffusion to occur at temperatures at which it would not be possible otherwise. Finally an account is given of low temperature creep of the type which does not involve recovery.

§ 1. EDGE AND SCREW DISLOCATIONS IN THE SIMPLE CUBIC LATTICE

It is the aim of this lecture to show the extent to which plastic flow, work hardening, recovery and creep can be explained in terms of the current theory of dislocations.

A dislocation is the name given to irregularities of certain types which are supposed to exist in crystalline solids. Although several accounts of their properties exist (for example Cottrell 1949), I shall begin this lecture by outlining some of them, in particular their ability to move only in definite directions relative to the crystal lattice. I would like to emphasize, moreover, that most of these properties do not depend on any particular assumptions about, for instance, the nature of the inter-atomic forces; they are the result, rather, of an analysis of the types of irregularity which are possible in crystals.

The most important property of a dislocation line is its Burgers vector; by this we mean the following. When a dislocation moves by the normal slip process, the material on one side of the plane swept out by the dislocation is displaced relative to that on the other by the Burgers vector. The two main types of dislocation are the edge and screw, the former with its Burgers vector perpendicular to the dislocation line, the latter with its Burgers vector parallel to it. For definitions and descriptions see Cottrell (1949, page 78). For our purpose we wish only to emphasize their different slip properties. These are illustrated in Figures 1 and 2 for a simple cubic lattice. In Figure 1 (*a*) the edge dislocation EF, with Burgers vector in the direction OB, can slip in the direction OB and will do so under the action of a stress. This is a rapid process; in a perfect crystal no activation by temperature is necessary. An edge dislocation can also move in the perpendicular direction OC, by generating or absorbing vacant lattice sites. This is a slow process, and needs an activation energy somewhat greater than that for self diffusion, as will be shown below.

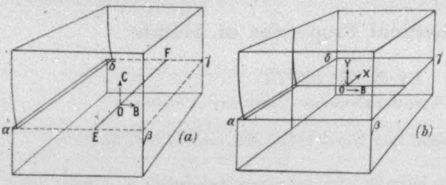


Figure 1. (*a*) An edge dislocation, and (*b*) a screw dislocation. In both cases OB is the Burgers vector. The edge dislocation can slip along OB, diffuse along OC; the screw dislocation can slip along OX, OY. $\alpha\beta\gamma\delta$ is a slip plane.

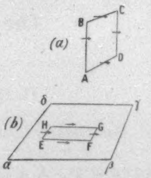


Figure 2. (*a*) A loop formed entirely of dislocations of edge type. (*b*) A dislocation loop in a slip plane $\alpha\beta\gamma\delta$. EF, GH are of screw type, FG, EH are edge. The arrows show the Burgers vector.

The screw dislocation shown in Figure 1 (*b*), on the other hand, by moving in the direction OX, can produce the same resultant slip as the edge of Figure 1 (*a*); it can also move in the direction OY, both processes taking place without activation energy. It cannot move through diffusion of vacant sites.

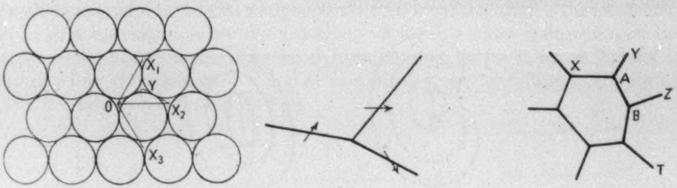
From edge and screw dislocations one can form 'dislocation loops' in two ways. (i) In a given slip plane, as in Figure 2 (*b*); such a loop is unstable, and will disappear by slip if left to itself, owing to the attraction between dislocations of opposite sign. In the presence of a stress σ, there is thus a critical size for which the loop will expand and not contract. Elementary calculations (Frank 1950) show that for the stresses σ normally applied to materials the activation energy necessary to form a loop which will expand is of order G^2b^3/σ, and if $\sigma/G \sim 1/100$ this amounts to several hundred electron volts. The production of loops by thermal agitation is therefore not possible. (ii) A loop can be formed of edge dislocations, as shown in Figure 2 (*a*). It will be seen that such a loop can move in the direction shown by the arrows. It represents either an extra plane of atoms or a plane of vacancies filling the area ABCD. Except by slipping right out of the crystal, it can disappear only by vacancy diffusion, which can remove the extra plane of atoms. Seitz (1950 a) has suggested that such rings are formed by the condensation of vacancies when a metal is cooled, and that this is an important source of dislocations in real materials.

§ 2. DISLOCATIONS IN CLOSE-PACKED STRUCTURES

The description of dislocations given above, valid for the simple cubic lattice, has to be modified in two important respects for close-packed structures.

Figure 3 (*a*) shows a close-packed plane of atoms. The figure shows that, in this plane, the Burgers vector can have three directions OX_1, OX_2 and OX_3, and that if two dislocations with vectors OX_1, OX_2 move across the plane the displacement is the same as if a dislocation with vector OX_3 had moved across it. Thus it is geometrically possible for three dislocations, with differing Burgers vectors, to meet in a point, as in Figure 3 (*b*). This leads to the hypothesis that real crystals may contain a network of dislocations (Figure 3 (*c*)), and, owing to the line energy or tension of dislocations, such an arrangement should be particularly stable. The network need not necessarily be confined to one plane.

The second way in which the closely-packed lattice differs from the simple cubic was first pointed out by Heidenreich and Shockley (1948). A dislocation in a close-packed plane, with Burgers vector say OX_1, can and will split into half-dislocations with Burgers vectors OY, YX_1. These two half-dislocations will repel each other; they will move apart until the energy of the 'stacking fault' so created balances the energy gained by moving them apart. The energy gained by this process is considerable. As a result, the dislocation lines of the network shown in Figure 3 (*c*) will, in general, be found in close-packed planes, for which their energy is much lower than for other planes. This is believed to be the reason why glide in close-packed structures shows a marked preference for close-packed planes, while for other structures (Fe, AgBr) various planes show glide.



(*a*) (*b*) (*c*)

Figure 3. (*a*) A close-packed plane, showing the Burgers vector for complete and for half dislocations. (*b*) Join of three dislocations. (*c*) Network of dislocations.

§ 3. THE MODEL OF FRANK AND READ FOR THE ORIGIN OF SLIP BANDS

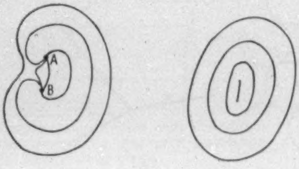
A first requisite of the theory is a model for the formation of slip bands. We have to explain why the deformation of a crystalline substance is not uniformly distributed through it (as is probably the case in amorphous materials), but is localized in slip bands. Observations with the electron microscope by Heidenreich and Shockley (1948) and by Brown (1949) have shown that in aluminium at any rate one observes deformation through about a thousand atomic distances on one plane.

The theory must thus provide a small number of fixed sources, where a large number of dislocation loops can be generated when a stress is applied, and also a mechanism for stopping slip on any plane containing a source when slip over about a thousand atomic distances has occurred.

A model of the nature of the sources was first given by Frank and Read (1950). These authors suppose that there exist in the crystal lengths of dislocation line which are securely anchored at the two ends. This may be a line such as AB of Figure 3 (*c*), lying in a close-packed plane, and held in position by other lines

AX, AY, BZ, BT, which lie in different planes. Or they may, as suggested by Seitz (1950 a), be elements such as AB of the loop of Figure 2 (*b*) formed by the condensation of vacancies.* Such lines will be spoken of as 'sources'.

Consider now what will happen if a stress is applied to a dislocation line AB, anchored at its extremities. The successive forms of the loop are shown in Figure 4 (*a*). It will be seen that ultimately a complete loop is formed, *without* the destruction of the original source. A Frank–Read source is thus capable of serving as the source of an infinite number of loops, each of which can expand over the whole slip plane. Thus, if a sufficiently large shear stress were applied to a crystal containing a single Frank–Read source, slip would occur to an indefinite extent along the plane in which the source lies. The model in the paper cited contains no mechanism for strain hardening.



(*a*) (*b*)

Figure 4. (*a*) Successive forms of a dislocation loop formed at a Frank–Read source; the plane of the paper is the slip plane. (*b*) Loops spreading out from a source.

In order that a source shall generate loops, a stress must be applied great enough to make the loop expand against its own line energy. If l is the length AB, Frank and Read show that a stress $\sigma=\alpha Gb/l$ is required. α is a numerical factor quite close to unity. G is the shear modulus and b the Burgers vector. Thus the longest sources will be the first to generate dislocations when an increasing stress is applied. To this stress must be added, of course, any internal stresses due to impurities, age hardening etc. (Mott and Nabarro 1948).

§ 4. EFFECT OF GRAIN BOUNDARIES

In an otherwise perfect crystal containing a single source, the loops will spread out until they come to the edge of the crystal. In a grain of a polycrystalline material, on the other hand, the loops will be held up at the grain boundary. If slip is to continue in the adjacent grains, starting on the line where the original dislocation comes to rest, the dislocations formed in the adjacent grains will normally be on planes of very high order. They will thus have much higher energy per unit length than the original dislocation.

Dislocation loops will thus pile up against the grain boundary. Now rows of dislocations in the same plane held up by a barrier have a number of properties of great interest in the theory. The first is that, given a row of n dislocations in a material acted on by a stress σ, the force on the leading dislocation is $nb\sigma$ per unit length, or n times the stress that would act on a single dislocation. In other words, a row of dislocations acts as a very effective stress raiser. Thus, as the dislocations pile up, a situation will be reached in which dislocation loops can be generated spontaneously in the next grain on the plane of high order mentioned above.

---

* In its state of lowest energy, a *small* loop would enclose a stacking fault in the plane ABCD, so that the dislocations would be sessile.

Another way of expressing the same thing is to say that a plane containing a row of dislocations shows as a whole practically no resistance to slip; the stress magnification at a distance r from the end is thus of order $\sqrt{(d/r)}$, where d is the linear dimension of the grain.*

§5. FORMATION OF VACANCIES BY MOVING DISLOCATIONS

Supposing one of the expanding dislocation loops of Figure 4 (*b*) cuts a screw dislocation, pictured as perpendicular to the plane of the paper. This will normally happen several times in the expansion of a loop in a real crystal.

The loop will then necessarily contain what we call a 'jog'†, i.e. a point where the dislocation jumps from one slip plane to an adjacent one.

Now a jog in an edge dislocation is equivalent to a small edge dislocation in another plane joining the two straight parts; thus it will not impede the movement of an edge dislocation in its slip direction. A jog in a screw dislocation, on the other hand, prevents the dislocation from moving unless it leaves behind it either a line of vacancies or a line of interstitial atoms according to the sign.‡ This is easily seen from Figure 5 (*b*).

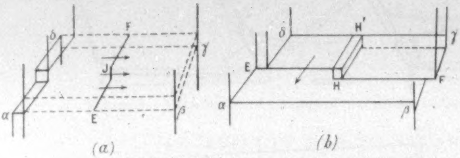


Figure 5. (*a*) A jog at J in an edge dislocation EF. (*b*) A jog H in a screw dislocation EF. The arrow shows the direction of movement. $\alpha\beta\gamma\delta$ is the slip plane.

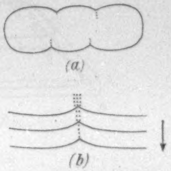


Figure 6. (*a*) Form of a dislocation loop which has crossed a number of screws. (*b*) Formation of vacancies or interstitial atoms by successive dislocations moving in the direction shown by arrow.

The first ring which spreads out from a Frank–Read source, then, will move without impediment in the direction of the Burgers vector; but in the perpendicular direction, where the dislocation has screw form, it will have to leave behind a row of vacanices or interstitial atoms whenever it crosses a screw dislocation. This will impede its motion; the ring will thus take up an oblong form (Figure 6 (*a*)).

We have, however, to consider whether the next ring will also be impeded; if not, the successive rings will soon push out the first one, so that it assumes a spherical form. Now if the temperature is low, so that the rows of vacancies or interstitial atoms remain stable until the next dislocation comes along, it will just add a second row of atoms (Figure 6 (*b*)). After three or four rows of atoms are added the work necessary to add more decreases rapidly; all one is doing is moving two unlike edge dislocations away from each other. If, however, the temperature is high enough and the rate of strain low enough for the vacancies (or interstitials) to diffuse away between the arrival of one dislocation and the next, the resistance to the motion of screw dislocations arriving after the first will be as great. Thus the tendency to form oblong loops should be especially marked at high temperatures.

---

* The mathematical development is given by Eshelby, Frank and Nabarro (1951).

† First pointed out by Thornton Read (unpublished).

‡ This was shown independently by F. C. Frank and by F. Seitz in work at present unpublished.

§ 6. CROSS SLIP

We now discuss what happens when two dislocation loops, generated on different planes, arrive at the stage where the screw component of one finds itself above and parallel to the screw component of the other, as shown in Figure 7(*a*). The screw components are of opposite sign, and therefore the lines PQ, P′Q′ attract each other. As we have stated in § 1, it is geometrically possible for them to move by slip out of the original plane of motion into one of the other slip planes. If this occurs, they will join up, leaving a single loop, as shown in Figure 7(*b*). In close-packed structures, where the dislocations are dissociated into two halves, it may however require considerable activation energy to move a screw from one plane to another, for along a certain length the two halves must be pushed together if this is to happen. If this is the case, one can achieve the same result by invoking the presence of other Frank–Read sources in the plane of cross slip, which are brought into operation by the stress round the piled-up screws of opposite sign.*

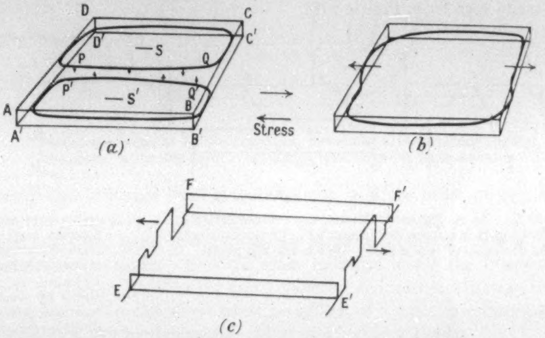


Figure 7. Showing formation of edge dislocations by cross slip. (*a*) two loops spreading out from sources S S′ in the planes ABCD, A′B′C′D′; (*b*) single dislocation loop after the screw dislocations PQ, P′Q′ have joined up by vertical displacement (cross slip); (*c*) two kinky edge dislocations left after a number of loops have joined up, and the screw dislocations moved to the edge of the crystal.

After several such loops have joined up, one will be left with two jagged edge dislocations (Figure 7(*c*)) joined up by two straight screws. Successive formation of loops by the original Frank–Read sources will lead to a succession of loops of this type. It should be noted that the horizontal parts of the edge dislocations follow one another in the same plane, but this is not necessarily the case for the vertical parts (cross slip).

We believe this mechanism to be responsible for the cross slip† observed for instance by Cahn (1951) in aluminium. Cahn has studied the slip bands on a single crystal of aluminium oriented as shown in Figure 8. The appearance of the slip bands was quite different on the top surface and the side surface.

---

* This suggestion is due to Dr. A. H. Cottrell (private communication).

† Cross slip has also been observed by Maddon, Mathewson and Hibbard (1948) in α-brass and by Ogilvie and Boas (1948) in aluminium. We do not wish to suggest that cross slip is necessarily *always* due to the mechanism outlined here.

Some of Cahn's photographs for the top surface are shown in Figures 9 and 10 (Plate *). The points to note are (*a*) the occurrence of cross slip, as suggested by the above theory, (*b*) that it is more marked at high temperatures, and (*c*) that, whereas the primary slip bands are sharp (as is expected for displacements due to successive dislocations from a given Frank–Read source), the cross slip is not.

On the side surface, on the other hand, no cross slip is to be expected, nor is it observed. What we do observe is deformation bands. These we believe to be the origin of strain hardening in metals which show them, and recovery and creep also have their origin there. These points will be discussed in § 8. We shall first, however, discuss certain relevant effects connected with the interaction between vacancies and dislocations.

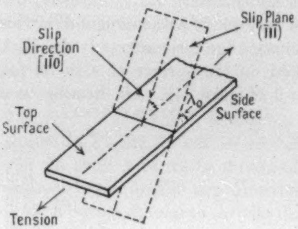


Figure 8. Orientation of crystals of aluminium in Cahn's experiments.

§ 7. EFFECTS DUE TO VACANCIES

We have already remarked that an edge dislocation can move perpendicularly to its slip plane by absorbing or giving off vacancies. It will be seen that vacancies can in fact only be generated at jogs, and the chance per unit time that a jog gives off a vacancy is $\nu e^{-W/\boldsymbol{k}T}$, where W is the energy for self-diffusion. If n/a is the number of jogs per unit length, it is easily seen that, under a force per unit length F, an edge dislocation will drift perpendicularly to its slip plane with velocity $(\nu a^4 F/\boldsymbol{k}T)(n/a)e^{-W/\boldsymbol{k}T}$. Apart from any jogs necessarily present in a ring that has cut screws during its formation, we may suppose that $n=e^{-U/\boldsymbol{k}T}$, where U is the energy required to form a jog. Thus the activation energy for this perpendicular motion of edge dislocation is $(W+U)$, rather larger than for self-diffusion.

In the same way, if the concentration of vacancies drops below the equilibrium value, the rate at which they are formed depends on T through the factor $e^{-(W+U)/\boldsymbol{k}T}$.

This temperature-dependent formation of vacancies must be contrasted sharply with the rows of vacancies that are forced into existence when a screw dislocation containing jogs is set in motion by an applied stress.

An effect due to the latter process is probably the increase in electrical resistance of metals, or at any rate a large part of it, which occurs on cold work. Vacancies would be very effective in increasing the electrical resistance. They would, however, be very mobile, and at comparatively low temperatures may be expected to diffuse to (edge) dislocations and disappear there. This accords with the observation that much of the electrical resistance due to cold work disappears on annealing at temperatures that scarcely affect the hardness. An example of this is the recent work of Molenaar and Aarts (1950) who have deformed polycrystalline Al, Cu and Ag at liquid air temperature, then annealed at room temperature for a few minutes. They find that on recooling to liquid air temperature nearly all the increased resistance has disappeared in Al, and some of it in Cu and Ag. In nickel the residual electrical resistance disappears at about 200° c., hardness, line width and magnetic induction at 600° c. (Wilson and Thomassen 1934).

---

* For Plates see end of issue.

As pointed out by Seitz (1950 b), the transient *increase* in electrical conductivity of alkali halides after cold work observed by Gyulai and Hartley (1928) must be ascribed to the same cause.

Perhaps the clearest example of a phenomenon in which vacancies are created and absorbed by stationary edge dislocations is the Kirkendall effect.* In this work markers are placed on the surface of a metal (e.g. CuZn), and another (e.g. Cu) plated on to it (Figure 11). On heating to a temperature at which diffusion is possible, the markers are observed to move inwards. One explains this by saying that the zinc diffuses more rapidly than the copper, and so diffuses out of the region AA′B′B. If, however, diffusion were due to the direct exchange of atoms, the diffusion coefficients of the two metals in the alloy at any point would have to be identical. One thus supposes that diffusion is due to the movement of vacancies, the copper and zinc atoms changing place only when adjacent to a vacancy.

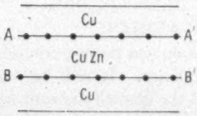


Figure 11. Arrangement for Kirkendall effect.

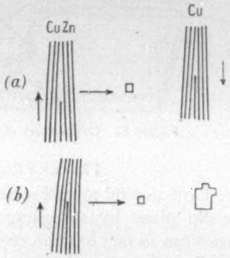


Figure 12. The mechanisms possible in the Kirkendall effect. The vertical arrows show the direction in which the dislocation moves, the horizontal arrows the direction of the drift of vacancies, marked with a square. In case (*a*) the vacancies condense on dislocations, in case (*b*) on macroscopic holes.

An elementary analysis of the Kirkendall effect may be made as follows: consider two adjacent planes of atoms distant a from each other, a being the lattice parameter. Let the proportion of lattice sites in the two planes occupied by zinc atoms be c, $c + a\partial c/\partial x$. Consider now a vacancy in either of these planes; we may suppose that it changes place more easily with a zinc than with a copper atom. Let the chance per unit time that it changes place with a zinc atom be P_A and with a copper atom P_B. The vacancy will then drift in the direction of excess copper concentration with velocity $(P_A - P_B)a^2\partial c/\partial x$. If the proportion of vacant sites is c_v (itself of course a function of c) and the atomic volume a^3 there is a mean drift of matter in the direction zinc to copper of amount $(P_A - P_B)c_v a^4 \partial c/\partial x$ cm^3/sec. It will be seen that the diffusion coefficient of the zinc relative to the lattice is $P_B c_v a^2$ and that of copper $P_A c_v a^2$; the mass drift thus depends on the inequality of the diffusion coefficients of the two constituents.

---

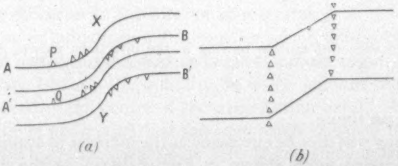
* For recent work cf. Correa and Mehl (1951).

This explanation, however, requires that sources or sinks of vacancies should be present within the grains of the alloy. The sources can only be dislocations (or grain boundaries), and the sinks also, unless the vacancies condense in the form of macroscopic holes. The presence of such holes has in fact been reported in certain cases. The two possibilities are illustrated in Figure 12.

An analysis by Seitz (1948) suggests that the actual density of dislocations in a material should be sufficient to replenish the vacancies at the rate required in phenomena of this kind.

§ 8. DEFORMATION BANDS.

Deformation bands shown in Figure 10 can be represented by dislocations as shown in Figure 13 (*a*). In some way, to be discussed below, a barrier XY is set up; the barrier can itself move in the slip direction; it thus straightens itself so as to take a planar form perpendicular to the slip direction. Dislocations then pile up against it from both sides.



(*a*) (*b*)

Figure 13. Representation of deformation bands by dislocations; (*a*) before and (*b*) after polygonization. The triangles represent the intersections with the plane of the paper of edge dislocations, the upright and inverted triangles representing dislocations of opposite sign. On the left are shown dislocations coming from sources at A, A′; on the right dislocations of opposite sign from B, B′.

Two questions arise: (*a*) What is the nature of the barrier? It can hardly be anything present in the crystal before deformation begins, at least in annealed single crystals. It must be formed by the deformation. (*b*) What stabilizes the deformation band? Why does it not disappear when the stress is taken off, the dislocations moving back to the original sources?

We have at present no fully satisfactory quantitative answer to either question, and the following suggestions are tentative. Consider what happens when an edge dislocation CDEF (Figure 14 (*a*)) coming from a pair of sources in the planes $\alpha\beta\gamma\delta$, $\alpha'\beta'\gamma'\delta'$ meets another dislocation AB of opposite sign, moving in the opposite direction. When they cross they form at once loops CDXA and BXEF, which can separate, and probably will do so, since they repel each other. Further movement can only occur by pulling out the segments DX, D′X′ as in Figure 14 (*b*), so that they acquire more and more a screw-like form. The barrier is then set up by the line energy of these cross links DX, D′X′.

A rather similar explanation may be attempted of the stability of the deformation band, when the applied stress is removed, against return of the outside dislocations, P, Q in Figure 13 (*a*), to the sources from which they came. If one had to do with straight edge dislocations, they would certainly be repelled; but, since the barrier is not uniform along a line perpendicular to the plane of the paper, it is possible that each dislocation P as it moves up will be deformed, and will take a form somewhat as in Figure 14 (*c*). Pieces of the dislocation will thus be forced close to the array BB′YX (Figure 13 (*a*)), which will exercise strong attraction on them.

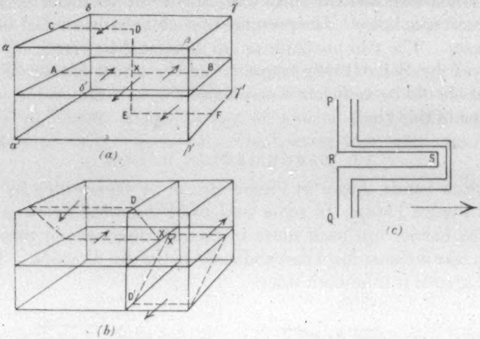


Figure 14. (*a*) Two edge dislocations, moving in the directions shown, meet at X. (*b*) The formation of two loops. (*c*) The pile-up of dislocations in a slip plane ($\alpha\beta\gamma\delta$). The segments RS are of screw type.

Work hardening will thus be envisaged as the increase of applied stress required to form the stabilizing loops shown in Figure 14 (*c*), as the deformation band becomes more and more tightly packed with dislocations.

§ 9. RECOVERY

Figure 10 (*b*) shows the observed polygonization of a slip plane, and Figure 13 (*b*) the representation in terms of dislocations. Two facts stand out: (*a*) that polygonization involves *essentially* the movement of edge dislocations out of their slip planes—i.e. a process involving diffusion of vacancies; (*b*) that the driving force is the repulsion between dislocations piled up in the same slip plane.

Various authors (e.g. Kuhlmann 1951) have considered recovery to be an escape of dislocations over a barrier; a treatment (Kuhlmann, Masing and Raffelsieper 1949, Cottrell and Aytekin 1950) is given somewhat as follows: the hardness or flow stress depends on (perhaps is proportional to) the number n of dislocations piled up against each barrier. The activation energy for the escape of a dislocation is taken to be of the form $W-\beta n$. The rate of change of n with time is thus given by the equation $dn/dt = A\exp\{-(W-\beta n)/\boldsymbol{k}T\}$ where A is a constant. Integration gives $n = n_0 - (\boldsymbol{k}T/\beta)\ln(1+t/t_0)$, where n_0 is the value of n at time $t=0$, and t_0 is given by $n_0 = -(\boldsymbol{k}T/\beta)\ln(\beta A t_0/\boldsymbol{k}T)$.

In the treatment given by Cottrell and Aytekin of steady state creep it is assumed (essentially) that σ, the flow stress, is proportional to n, so that we may write $d\sigma/dt = A' \exp\{-(W-b\sigma)/\boldsymbol{k}T\}$, and $\sigma = \sigma_0 - (\boldsymbol{k}T/b)\ln\{1+t/t_0\}$. Then, if h is the rate of hardening ($h = d\sigma/d\gamma$, where γ is the strain) and r the above rate ($d\sigma/dt$) of recovery, the creep rate $\kappa(=d\gamma/dt)$ is $\kappa = r/h = (A'/h)\exp\{-(W-b\sigma)/\boldsymbol{k}T\}$. Cottrell and Aytekin (1950) thus relate the values of b for creep and recovery. For zinc they obtain values of the order 200 cal/gm. mol per gm/mm^2; for the stresses applied (of order 50 gm/mm^2), $b\sigma$ is of order 10,000 gm/mol.

Kuhlmann (1951), for instance, believes $W-b\sigma$ to the the activation energy required for a dislocation to escape by a slip process over a barrier, the appropriate number of dislocations being piled up behind. The discussion of Mott and Nabarro (1948) shows, however, that energies of this sort are in general very large ($W>10$ ev.) unless σ is nearly great enough for spontaneous escape without the help of temperature (see also Cottrell and Aytekin 1950, p. 412). It seems to us much more likely that in recovery and steady state creep we have *always* to do with a process akin to polygonization, in which dislocations move out of their slip planes. W will then be, as already shown, slightly greater than the energy for self diffusion; this assumption will certainly give energies of the right order for the interpretation of creep. It also explains why:

(*a*) polycrystalline metals creep faster than single crystals. In fine grained materials the dislocations may pile up on the grain boundaries instead of, or as well as, in deformation bands. Diffusion of atoms along grain boundaries can then remove dislocations from the slip bands. This is believed to be a more rapid process than diffusion in the interior of a crystal.

(*b*) The addition of substitutional impurities which diffuse rapidly, and thus increase the rate of self-diffusion, may increase the creep rate.

If these ideas are correct, the quantity $b\sigma$ is the amount that the activation energy W is decreased in the centre of the deformation band. If σ_0 is the stress there, this should be of order $\sigma_0 a^3$. If this is to reach values of the order 10 kcal. in zinc, for instance, σ_0 must itself reach very large values indeed, of the order $0{\cdot}1G$, where G is the shear modulus. The model in which several hundred dislocations pile up behind each other makes this possible; if n is the number in a given slip plane, then for applied stresses σ we expect $\sigma_0 = n\sigma$. These very large stresses are essential to any diffusion theory of recovery, as they are also to the explanation of asterisms.

Transient creep following equations of the type $\gamma = \text{const.}\, t^{1/3}$ is probably to be explained by a similar mechanism: dislocations are squeezed out of the slip planes, at a gradually decreasing rate as the material between them fills up with dislocations.

§ 10. HEXAGONAL METALS

The very tentative discussion of work hardening given here applies essentially to cubic metals, in which cross slip is possible. It is the cross slip which, we believe, makes it possible for edge dislocations moving in one direction to hold up those moving in the opposition direction, and so form deformation bands.

In hexagonal metals cross slip has not been observed, and we should not expect it according to the mechanism outlined here. Deformation bands and asterisms in the Laue diagrams are not observed either. Hardening must here be by a different mechanism. Possibly, as suggested by Seitz (1951), the vacancies formed when dislocations move will eventually bring slip to rest in the plane of a given Frank–Read source, even if no deformation bands are formed.

A *slow* hardening mechanism due to some such cause as this may be operative for cubic metals also, if they are stressed under conditions such that deformation bands are not formed. To form deformation bands one requires that not too low a density of Frank–Read sources should be operative *at the same time*. A good single crystal, uniformly stressed under conditions such that dislocations could escape from the surface, should not necessarily form deformation bands, but should harden like a hexagonal crystal. This may account for an observation of Andrade (1951), that silver crystals stressed in argon harden much more slowly than in air. The oxide film, by preventing the exit of dislocations, sets up inhomogeneous stresses which start the formation of the deformation band.

§ 11. LOW-TEMPERATURE CREEP

There is, however, one type of creep in which movement of vacancies probably does not occur. This is the logarithmic creep observed in metals and alloys at low temperatures (Davis and Thompson 1950). A theory of this in terms of an exhaustion mechanism was given by Mott and Nabarro (1948). In creep one always has to consider that a certain amount of instantaneous deformation takes place as soon as the stress is applied. If we assume that deformation bands are formed, the subsequent creep represents the movement of the dislocations into the deformation bands under the influence of stress and temperatures.

Until this process is better understood, it does not seem worth while to discuss in detail its connection with the exhaustion mechanism. We shall show, however, that a theory of the type given by Orowan (1947) leads essentially to the same formula, though with different values of the constants.

We assume, then, that low temperature creep of this kind is due to temperature-activated slip in slip bands already formed. Thus, in such slip bands, we have an array of, say, n dislocations pressing against the deformation bands. If l is the distance between slip bands, and p the chance per unit time that a dislocation moves into a deformation band, the creep rate is $d\epsilon/dt = bp/l$.

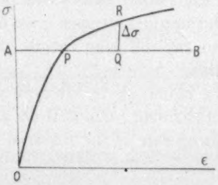


Figure 15. Stress–strain curve.

Consider now a substance with a stress–strain curve such as that shown in Figure 15. A stress σ equal to OA is applied, so that an instantaneous extension AP occurs. At any subsequent time let the extension be AQ. Then on our model, assuming that hardening takes place as rapidly during creep as in more rapid extensions, QR represents the stress needed to cause flow without the help of temperature. We need to know the activation energy required for slip to take place without an increase in stress. Denoting QR by $\Delta\sigma$, Orowan takes this to be $\Delta\sigma^2 V/2G$, where V is the minimum volume in which a thermal fluctuation of stress must occur for slip to take place. An analysis by Mott and Nabarro (1948) gives instead $0{\cdot}15b\lambda^2\sigma(\Delta\sigma/\sigma)^{2/3}$, where λ is a length of order 10^{-6} cm. We shall set for the activation energy $\Omega\sigma(\Delta\sigma/\sigma)^m$; on any analysis $\Omega\sigma$ should be large compared with Gb^3. For p, then, we may set

$$p = \nu \exp\{-\Omega\sigma(\Delta\sigma/\sigma)^m/\boldsymbol{k}T\}.$$

Also, treating hardening in the range PR as linear, and writing ϵ' for the extension PQ during creep, we may write $\Delta\sigma/\sigma = r\epsilon'$ where $r = d(\ln\sigma)/d\epsilon$. Then $d\epsilon'/dt = \nu' \exp\{-\mu\epsilon'^m\}$, where $\nu' = b\nu/l$ and $\mu = \Omega\sigma r^m/\boldsymbol{k}T \gg 1$. Since ν' may be of order $10^5\,\mathrm{sec}^{-1}$, $\mu\epsilon'^m$ will be large when creep has slowed down.

Integration by parts gives $\{\exp(\mu\epsilon'^m)\epsilon'^{1-m}/\mu m\} +$ terms in $\mu^{-2} = \nu' t$, or $\epsilon' = A\{\ln\gamma t\}^{1/m}$ where $A = \mu^{1/m}$, $\gamma = \mu m \nu' \epsilon'^{m-1}$. In view of the very large range over which t varies in most experiments and the small range of ϵ', γ may be considered sensibly constant; the formula is thus of just the same type as that deducible from the exhaustion hypothesis (Mott and Nabarro 1948).

REFERENCES

Andrade, E. N. da C., 1951, *Proc. Roy. Soc.* A, in the press.

Brown, A. F., 1949, *Nature, Lond.*, **161,** 961.

Cahn, R. W., 1951, *J. Inst. Met.*, **18,** 129.

Correa, L. C., and Mehl, R. F., 1951, *J. Metals* (*Metals Trans.*), **191,** 155.

Cottrell, A. H., 1949, *Theory of Dislocations*; *Progress in Metal Physics*, Vol. 1 (London: Butterworth's Scientific Publications), p. 77.

Cottrell, A. H., and Aytekin, V., 1950, *J. Inst. Met.*, **77,** 389.

Davis, M., and Thompson, N., 1950, *Proc. Phys. Soc.* B, **63,** 847.

Eshelby, J. D., Frank, F. C., and Nabarro, F. R. S., 1951, *Phil. Mag.*, **42,** 351.

Frank, F. C., 1950, *Report of Pittsburgh Conference on Plastic Deformation of Crystals* (Washington: Carnegie Institute of Technology and Office of Naval Research).

Frank, F. C., and Read, W. T., 1950, *Phys. Rev.*, **79,** 722.

Gyulai, Z., and Hartley, D., 1928, *Z. Phys.*, **51,** 378.

Heidenreich, R. D., and Shockley, W., 1948, *The Strength of Solids* (London: Physical Society), p. 57.

Kuhlmann, D., 1951, *Proc. Phys. Soc.* A, **64,** 140.

Kuhlmann, D., Masing, G., and Raffelsieper, J., 1949, *Z. Metallkunde*, **40,** 241.

Maddon, R., Mathewson, C. H., and Hibbard, W. R., 1948, *Trans. Amer. Inst. Min. Metall. Engrs.*, **175,** 86; 1950, *Ibid.*, **185,** 527.

Molenaar, J., and Aarts, W. H., 1950, *Nature, Lond.*, **166,** 690.

Mott, N. F., and Nabarro, F. R. N., 1948, *The Strength of Solids* (London: Physical Society), p. 11.

Ogilvie, G. J., and Boas, W., 1948, *Trans. Amer. Inst. Min. Metall. Engrs.*, **175,** 102.

Orowan, E., 1947, *West of Scotland Iron and Steel Inst.*, 45.

Seitz, F, 1948, *Phys. Rev.*, **74,** 1505; 1950 a, *Ibid.*, **79,** 723, 890, 1002, 1003; 1950 b, *Ibid.*, **80,** 239; 1951, *Phil. Mag.*, in the press.

Wilson, J. E., and Thomassen, L., 1934, *Trans. Amer. Soc. Metals*, **22,** 769.

The Deformation and Ageing of Mild Steel: II Characteristics of the Lüders Deformation

By E. O. HALL
Cavendish Laboratory, Cambridge

Communicated by W. L. Bragg; MS. received 8*th March* 1951

ABSTRACT. In this paper, some of the factors which influence the appearance of Lüders bands in mild steel are studied. It is shown that the Lüders band is adequately described by a uniform shear front, spreading over the specimen. In coarse-grained specimens, experiments indicate that this front becomes diffuse; diffuse bands are also present in strain-aged material, but here the diffuse fronts become sharper as the ageing becomes progressively longer.

§ 1. INTRODUCTION

In a previous paper (Sylwestrowicz and Hall 1951), hereafter referred to as I, work was described on the propagation of Lüders bands in mild steel. In this present paper, further experiments will be described on the appearance of Lüders bands in thin strip and wire specimens, in an attempt to elucidate some of the factors which influence their form.

§ 2. THE FORM OF A LÜDERS BAND

The patterns which Lüders bands may present on the surface of a polished test-piece are many and varied. Certain stamped tin-plate articles, such as pie and cake dishes, often show very complicated patterns of 'stretcher strains'. Under simpler stress conditions, however, as in a normal tensile test, the bands can still present a very complex appearance. Figure 1 (Plate I*), taken in this laboratory by Mr. W. M. Lomer, shows an intricate pattern of intersecting fronts. The upper part of the specimen is undeformed, and a severe roughening of the polished surface is seen in the deformed areas. Under these conditions, the lower yield stress fluctuates rapidly. Most published pictures of Lüders bands fall into this category: it is necessary to have thin strip or wire specimens before single bands may be obtained.

Figure 2 (Plate I) shows a typical simple band, observed in a strip of mild steel $\frac{1}{2}$ in. × $\frac{1}{16}$ in. The band has again spread from below, so that the upper part of the specimen is still undeformed. The faint parallel striations in the deformed area are the result of periodically stopping and starting the test; the various strain rates that result give correspondingly varying strains (see I), which show up as these faint striations. Under constant strain rate, the lower yield stress is very uniform.

On examining any such specimen, it is easily seen that the deformed and undeformed areas are no longer coplanar, the disorientation being of the order of 1°. Consequently, the boundary of the band is easily seen under any Schlieren illumination method which is sensitive to small angular changes. These photographs show that the specimen has kinked in the wide front face of the specimen. However, an angular change has also occurred in the plane of the thin side face.

---

* For Plates see end of issue.

This is seen by scribing lines on the specimen before the test, and Figure 3 (Plate I) shows one such specimen. Here the test has been stopped before the two bands have completely run together, leaving a small triangular area undeformed, and looking along the lines a distinct angular deviation can be seen at the edges of the band, again of the order of 1°.

Thus, the form of the test-piece during a tensile test must be as shown in Figure 4 (*a*). The kink in both the side and front faces causes the front of the band to lie close to the plane of maximum shear, and is not, as the photographs of the front face would seem to indicate, at some 56° to the tension axis.

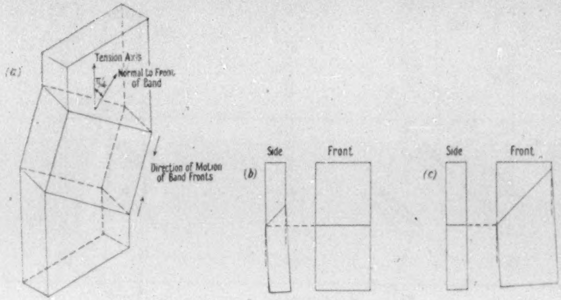


Figure 4. Diagrammatic view of test piece.

Since the width of the strips used in these experiments was eight times the thickness, it might be expected that the appearance of the band would be more as shown in Figure 4 (*b*), for the specimen can accommodate the bending moment set up at the band edge more easily in this position. However, the configuration of the band tends more towards Figure 4 (*c*). The explanation for this may lie in the question of the grip constraint. With the testing machine used here, the top grip could swing in the plane of the front face of the specimen, and was to that extent self-aligning. Consequently, the specimen could accommodate the bend in the thin face more readily, and bands of the type shown in Figure 2 were formed, a compromise between the extreme cases shown in Figures 4 (*b*) and 4 (*c*).

The constraint of the grips must be one important factor in determining whether bands are simple or complex in appearance, although other factors, such as alignment, heat-treatment, and resulting grain size may also play their parts.

§ 3. THE UNIFORM SHEAR MECHANISM

Since Lüders bands form with their fronts around 45° to the tension axis, it has been suggested before that the mechanism of propagation was a uniform shear front, moving along the specimen, but because of the apparent difficulty of obtaining simple bands, this hypothesis has only been tentative. However, it is felt that the evidence here supports this conclusion.

The case of a uniform shear front in motion is shown in Figure 5. As each elementary section at the front of the band yields, it is sheared over under the applied stress, so that if the specimen were free of constraints, it would appear, in exaggerated form, as in Figure 5 (*a*). However, this shear displacement is naturally straightened immediately by the constraints of the grips, by a simple bend about a line in the plane of the band front. At the same time, the factor impeding normal slip has been destroyed. The material then flows under the lower yield stress, that is, it creeps. As creep continues, the local stress concentrations build up, until the next lamella yields, and the deformation is repeated. The final shape of the specimen is then as shown in Figure 5 (*b*).

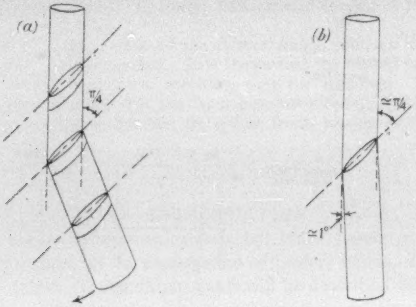


Figure 5. The uniform shear mechanism.

This combination of shear, bend and flow is in many ways comparable with that of flexural glide (Mark, Polanyi and Schmid 1922) observed in metal single crystals, with the exception that there is a sharp front to the deformation. This mode of deformation also explains the elliptical cross section of wire specimens after the Lüders deformation (see I). As each section is sheared over and pulled straight by the load, the cross section of the wire normal to the tension axis becomes elliptical, but the major axis of the ellipse is still equal to the diameter of the wire. The flow that follows decreases both axes, so that the resulting profile is that mentioned in I. This peculiar mode of deformation may cause the deviation of the band front from 45°, but further investigations are being carried out here.

§ 4. DIFFUSE LÜDERS BANDS

Microscopic examination of Lüders bands indicates that, in fine-grained material, the boundary of the deformation is extremely sharp, perhaps even to a grain. During tests on coarse-grained specimens with a special optical device, the writer (Hall 1950) noted a change in the form of the records, which indicated that a modified form of deformation was occurring. However, before describing the records, it will be convenient here to outline again the optical method used, so that records can more readily be appreciated.

A beam of light from a Western Electric 'concentrated arc' lamp is focused on a point about half-way down the wire tensile specimen, and the image of this area is then focused on a piece of recording paper in a rotating drum camera, giving an optical magnification of 15 to 50 times. Drawing marks, scratches and flecks of rust give a spotted optical image, so that as the camera rotates, a series of lines is recorded on the photographic paper. From the traces obtained during the test, the mode of deformation may be determined.

Typical records are shown diagrammatically in Figure 6. In the extensometer, the bottom end of the specimen is pulled down, so that the traces on the record are the inverted image of the motion of the centre of the wire. The start of each trace is under zero stress. As the load comes on, the centre moves down with a velocity $V/2$. When a Lüders band is nucleated at the bottom grip, as shown in Figure 6 (*a*), no deformation occurs outside the band, the centre remains fixed in space, and the traces are horizontal. As the edge of the band moves through the illuminated area, each point on the wire suddenly moves downwards with the full velocity V of the straining head as the band passes. Thus, the gradient of the trace is twice that in the elastic region. The line of transition is staggered across the specimen, as shown by the hypotenuse of the dotted triangle in the diagram. When the band has reached the top grip, deformation is from then on homogeneous, and the gradient of the trace is the same as in the elastic region.

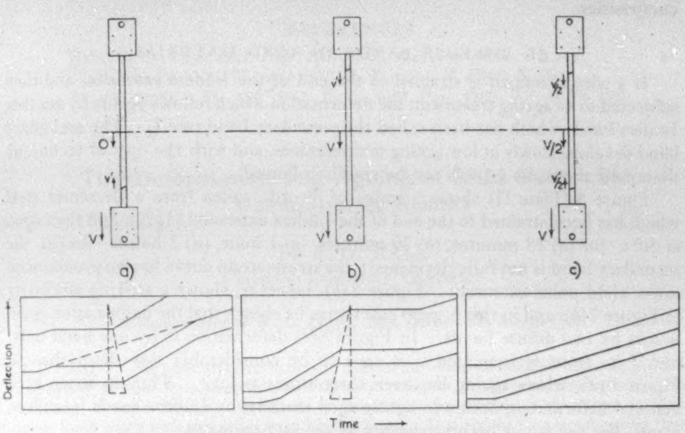


Figure 6. Diagrammatic optical traces.

Figure 6 (*b*) shows the traces in the similar case when the band is nucleated in the top grip, while Figure 6 (*c*) shows the case of two bands generated simultaneously.

Figure 7 (*a*) (Plate II) now shows an actual trace from a specimen of Bessemer steel, with 135 grains/mm. It shows that the band edge, as it passed the centre of the wire, had a very sharp front—the gradients of the traces change abruptly. An interesting contrast is shown by Figure 7 (*d*). Here, the gradients of the traces alter slowly as the band passes, swinging over to a horizontal position in the space of a minute. Thus, the edge of the band has lost its sharp kink, and can no longer be seen optically. This record is taken from PXQ iron, with 24 grains/mm.

Figures 7 (*b*) and 7 (*c*) appear to be somewhat different, but by marking gauge lengths on the specimen, and photographing it at various stages of the deformation, it may be shown that the deformation is occurring by two diffuse bands, one starting from each grip. The resemblance to Figure 6 (*c*) is now clear. Figure 7 (*b*) is taken from Armco iron (18 grains/mm.) and 7 (*b*) from the Bessemer steel (27 grains/mm.). The phenomenon is thus quite general in coarse-grained mild steel.

The Lüders bands appear to lose their sharp front at some critical grain size, but the exact value is rather difficult to determine. Work in this laboratory indicates that it may lie between 30 and 40 grains/mm., and is independent of composition.

§ 5. DIFFUSE BANDS IN AGED MATERIAL

If a wire specimen is strained to the end of the Lüders extension and then subjected to an ageing treatment, the deformation which follows occurs by another Lüders band, which has been called the secondary band (see I). The secondary band develops slowly at low ageing temperatures, and with the optical technique described above, its growth can be readily followed.

Figure 8 (Plate II) shows a series of records, taken from a Bessemer steel which has been strained to the end of the Lüders extension (3½%), and then aged at 50° c. for (*a*) 15 minutes, (*b*) 30 minutes, (*c*) 1 hour, (*d*) 3 hours. In (*a*), the secondary band is not fully developed; the stress–strain curve has no pronounced lower yield point extension. Figure 8 (*b*), however, shows a striking similarity to Figure 7 (*b*), and in this present case it may be shown that the deformation again occurs by two diffuse bands. In Figure 8 (*c*) deformation is by one band only, but if its front is examined it is seen to be considerably less sharp than in Figure 8 (*d*), where ageing has been three times as long. Thus in strain-aged material deformation occurs in lightly aged material by Lüders bands by diffuse fronts, but the fronts become sharper as ageing progresses.

As ageing proceeds, and the band becomes sharper, so also does the propagation stress for the secondary band rise. If the difference between σ'_{LYP}, the propagation stress for the secondary band, and σ_{LYP}, the propagation stress for the primary band, are plotted as a function of ageing time and ageing temperature, a series of curves are obtained very similar to the set given in Figure 3 of I. From a study of the stress–strain curves and the optical records, then, it is possible to divide the secondary Lüders deformations into three rough categories as shown in the Table. This Table does not mean that the form of the band determines the value of the propagation stress, for other variables may well be present, but the values of $\sigma'_{\mathrm{LYP}} - \sigma_{\mathrm{LYP}}$ given here seem to hold over a wide range of grain sizes.

Classification of Secondary Lüders Bands

$\sigma'_{\mathrm{LYP}} - \sigma_{\mathrm{LYP}}$ (kg/mm²)	Type of deformation
<1	Bands not fully formed.
1–4·5	Diffuse bands present.
$>4{\cdot}5$	Sharp bands present.

Both diffuse and sharp bands can propagate under constant stress. However, if the material is so coarse-grained that the primary Lüders band is diffuse, then no amount of straining and ageing will ever cause the secondary band to become sharp.

All these effects find a ready interpretation in terms of a grain-boundary theory of the deformation which is discussed in the following paper (Hall 1951).

ACKNOWLEDGMENTS

The writer wishes to thank Professor Sir Lawrence Bragg and Dr. E. Orowan for their interest in this work, and Mr. W. M. Lomer for supplying him with Figure 1 of the paper. Scholarships from the Royal Commission for the Exhibition of 1851, and the Lord Rutherford Memorial Fellowship have enabled this research to take place; it is being extended under the auspices of the British Iron and Steel Research Association.

REFERENCES

Hall, E. O., 1950, *Proc. Phys. Soc.* B, **63**, 724; 1951, *Ibid.*, **64**, 747.
Mark, H., Polanyi, M., and Schmid, E., 1922, *Z. Phys.*, **5**, 58.
Sylwestrowicz, W., and Hall, E. O., 1951, *Proc. Phys. Soc.* B, **64**, 495.

The Deformation and Ageing of Mild Steel: III Discussion of Results

By E. O. HALL
Cavendish Laboratory, Cambridge

Communicated by W. L. Bragg; MS. received 8*th March* 1951

ABSTRACT. An attempt is made here to explain the observed phenomena in the yielding and ageing of mild steel, described in two previous papers, in the general terms of a grain-boundary theory. On this hypothesis, a satisfactory explanation of the variation of the lower yield point with grain size may be developed. It is shown that strain-ageing must involve two processes: a healing of the grain-boundary films, coupled with a hardening in the grains themselves. A discussion of the possible nature of the grain-boundary film is also undertaken.

§1. INTRODUCTION

In two previous papers (Sylwestrowicz and Hall 1951, Hall 1951, which will be referred to as I and II) an experimental study has been made of some of the factors which influence the behaviour of mild steel in tensile tests. In this paper, a little more experimental evidence will be supplied, and the results correlated qualitatively in terms of a grain-boundary theory.

§2. THE UPPER YIELD POINT

There is now no doubt, from the work of Fast (1950) and others, that small amounts of carbon and nitrogen present in a matrix of ferrite will cause the yield point and ageing phenomena found in mild steel. In I, a comparison was made between the two current theories of the upper yield point, the old grain-boundary theory of Nadai (1924) and Kuroda (1938), and the more recent 'dislocation clogging' mechanism of Cottrell and Bilby (1949). Without identifying any particular component at the grain-boundary, it was suggested in that paper that some form of grain-boundary theory was more likely to explain the observed results.

The one objection to the grain-boundary theory most frequently heard (Cottrell 1948, Bilby 1950) is that once the matrix of ferrite has yielded, the grain boundary films alone have to withstand the load. Arguing from an analogy of a series of hard, brittle wires and soft ductile ones in parallel array pulled together in tension, Bilby (1950) has shown that a model with correct proportions of these brittle and plastic elements would have an extremely small upper yield point. In the opinion of the present writer, however, this analogy is incorrect, and it would be wise to examine the 'honeycomb' hypothesis of Kuroda in the light of our present knowledge of dislocation theory.

It is known that if a dislocation in a matrix approaches a region, such as a grain-boundary film, which has a higher yield stress than the matrix itself, then this region constitutes an effective potential barrier for the passage of the dislocation. In other words, the dislocation will experience a repulsion near the grain boundary interface, and dislocations will pile up along the glide plane behind the grain boundary film, until the stress concentrations around the tip of the slip band cause the film to yield. Deformation is then transferred to the next grain.

Cottrell (1949) has shown that the force on such an obstacle from an array of n dislocations of like sign is given by $P=na\sigma$ where a is the interatomic spacing, and σ the applied shear stress, or, in other words, the dislocation array produces a local stress concentration in the vicinity of the obstacle of n times the external stress.

This stress concentration which the boundary must withstand is very much less than in Bilby's model, where it is assumed that all the load is being carried by the thin boundary films. A direct comparison of the two models may be made if we let L be the length of the slip plane, l the spacing of the dislocations, and b the thickness of the boundary film. Then if after the ferrite yields, it carries no stress, the stress concentration factor in the film is L/b. In the model based on an array of dislocations, the stress concentration is n where $nl=L$. For a grain diameter of 10^{-3} cm. and a maximum thickness for the film of ten atomic spacings, $L/b \simeq 10^4$. In the other case, we must assume a dislocation density of 10^8 dislocation lines/cm^2 in the annealed ferrite, or an average of ten dislocation lines per slip plane in material of the same grain size. Provided no multiplication of the dislocations takes place until the boundary film yields, this model gives a stress concentration factor about a thousand times smaller than in the model composed of ductile and brittle wires, and a more reasonable value for the physical strength of the boundary.

The exact nature or the strength of the boundary films must still remain unknown; some of its necessary properties will be examined in the final paragraph of this paper. The concept of dislocation arrays will be extended further in the next section on the lower yield point.

In the first paper in this series, it has been suggested that a theory of this nature would be necessary to explain the observed dependence of the upper yield point on grain size. It is interesting to see how this theory, coupled with the 'nucleation' hypothesis advanced in the first paper, gives the correct qualitative grain-size dependence. The larger the grains, the more rapidly is the size of the plastically deformed nucleus increased by the breakdown of a single boundary, and the more rapidly is the state reached when the nucleus will spread to form the first Lüders band. Thus, one would expect the difference $\sigma_{\mathrm{UYP}}-\sigma_{\mathrm{LYP}}$ to be smaller the larger the crystal size, and the results in I (Table 2) confirm this.

§ 3. THE LOWER YIELD POINT

If this concept of the grain-boundary theory is correct, then it is possible to explain many of the facts about the lower yield point, and in particular, its variation with grain size.

The relationship between the lower yield point and grain size has been determined experimentally by the writer, using the three materials referred to in I, and the results are displayed in Figure 1. It is seen that all the points, with the exception of the coarsest-grained Bessemer steel, lie on a single curve. The Bessemer steel, however, contains a much larger amount of nitrogen than the other two, and a high level of impurity, but it is interesting that the Armco and open-hearth steels, containing 0·02% and 0·06% carbon respectively, have identical values of the lower yield point within the limits of experimental error. The full curve in this Figure lies parallel to and a little below the curve of results of Winlock and Leiter (1937); their results, however, covered a smaller range of grain sizes, and were taken at much higher strain rates (0·002/min., instead of 10^{-4}/min. as in the present case). Thus, higher values of the lower yield point would be expected in their case.

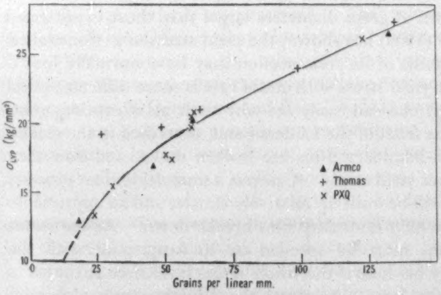


Figure 1.

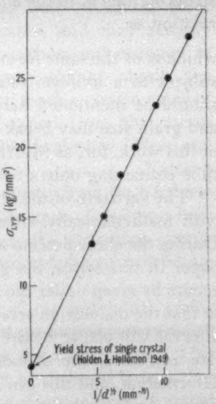


Figure 2.

If instead of Figure 1, values of σ_{LYP} from the full curve are plotted against the inverse square root of the grain diameter d, it is found that the relation beween the two quantities is linear (Figure 2); the graph cuts the y axis at the value for the yield stress of single crystals determined by Holden and Hollomon (1949 a). Thus

$$\sigma_{LYP} - \sigma' \propto 1/d^{1/2} \qquad \ldots\ldots(1)$$

where σ' is the yield stress for a single crystal.

This relationship follows from theoretical calculations by Eshelby, Frank and Nabarro (1951). The model used in their work is outlined in the following paragraph.

The stress at a distance r from a dislocation is of the form A/r, where A is a constant. Let there be a succession of n of these dislocations of like sign, forced in the distance of increasing x by a uniform applied shear stress σ_0, while the leading dislocation is held at $x=0$ by an obstacle, such as a grain-boundary film, which exerts no long-range forces. Equilibrium is then established between applied force, the obstacle and the mutual repulsion of the dislocations. It may then be shown that, to a close approximation, the distance between the leading and the second dislocation is $d' \simeq 2A/n\sigma_0$ while the distance between the first and last dislocation is $L \simeq 2An/\sigma_0$.

The additional stress σ at a point on the glide plane a distance x ahead of the first dislocation, such that $d'/15 \ll x \ll L$ is given by

$$\sigma/\sigma_0 \simeq (L/x)^{1/2}. \qquad \ldots\ldots(2)$$

This relationship is exactly analogous to the stress concentration at a distance x from the end of an internal crack of the type postulated by Griffiths (1920) to explain the low brittle strength of solids.

If we suppose that the grain-boundary film will break down when the additional stress σ reaches a certain critical value, then the deformation will be transmitted to the next grain. This state will occur when the end dislocation is at a fixed distance x from the boundary: furthermore, we may identify L, the length of the slip band, with the mean grain diameter d. Thus, rearranging, we find that $\sigma_0 \propto 1/d^{1/2}$.

However, since flow is impossible in a single crystal at stresses below its yield stress σ', then σ_0 in the analysis should be replaced by $\sigma_0 - \sigma'$ giving the grain-size relation as

$$\sigma_0 - \sigma' \propto 1/d^{1/2}, \qquad \ldots\ldots(3)$$

which is of the same form as (1). This final step implies that each dislocation is subject to a uniform retarding force σ' in its passage through each grain. It should be mentioned here that this relationship between the lower yield stress and grain size may break down at grain diameters larger than those investigated in this work, for, as Morrison (1939) has shown, the yield stress of a thin-walled tube containing only a few grains in its cross section may be abnormally low.

The variation of the lower yield stress with strain rate is more difficult to deal with mathematically, although one can easily see why such an effect does exist. During the shear process at the front of the Lüders band, described in the second paper in this series, the grain-boundary films are broken down, and flow then occurs by creep under the lower yield stress. Creep is a time dependent process, so that the dislocation arrays will be built up relatively slowly, and an appreciable interval will elapse before the next boundary film breaks down. As the strain rate increases, the dislocations must be speeded up to keep pace with the deformation, and this can only be done if the whole stress level rises at the same time. And as the propagation stress increases, the Lüders strain also rises (see I).

The grain-boundary theory also explains the appearance of diffuse Lüders bands in coarse-grained material (see II). Here the individual grains are so large that part of the front of deformation may get well ahead of another part which is held up, say, near a patch of smaller included grains. Then the sharp shear kink disappears, and we have the phenomenon of a typical diffuse band.

§ 4. STRAIN-AGEING PHENOMENA

When the Lüders band covers the entire gauge length of the specimen, all the grain boundary films are broken down and weakened, and the material is able to deform homogeneously. During the ageing process, the films begin to heal around the strain-hardened grains by the diffusion of carbon and nitrogen, with a resulting increase in their yield stress and the yield stress of the whole specimen, and deformation must again occur by a Lüders band—the secondary band. In lightly aged material, however, the boundaries have not regained their full strength, and the effect is the same as in material where the grain boundaries are far apart. Deformation is by a diffuse band. As the ageing becomes more advanced, more and more boundaries become restored to their original state, the front of the band is straightened and becomes sharp.

This analysis of course only applies to material which was sufficiently fine-grained to deform by a sharp band in the virgin state. If the grain size is so large that the initial deformation was by a diffuse band, then naturally no amount of straining and ageing will bring the grain boundaries any closer together, and deformation will always be by diffuse bands.

Thus, as the grain-boundary films reform around the strain-hardened material, the yield stress of the bulk material rises, and a set of curves similar to Figure 3 of I is obtained. However, it is not possible, from a set of curves such as this, to calculate the activation energy for the ageing process, because two processes are at work. In addition to the healing of the boundary films, Holden and Hollomon (1949 b) and Cottrell and Churchman (1949) have also shown that single crystals may strain-age—consequently we must consider the individual grains of the matrix as hardening after the deformation as well. Each of these processes will raise the yield stress by different amounts, and may actually occur at different rates, for the diffusion constants of carbon and nitrogen in α-iron may be structure sensitive. Consequently, one cannot obtain the activation energy simply by measuring the time needed to reach a certain value of $\sigma_{\mathrm{UYP}}-\sigma_{\mathrm{LYP}}$ at various temperatures. The values obtained from the curves in I are all much lower than the accepted value of 19·8 kcal/mole (Wert and Zener 1949), but tend to this value at higher values of $\sigma_{\mathrm{UYP}}-\sigma_{\mathrm{LYP}}$. Nabarro (1948) was able to calculate this value for the activation energy from the hardness measurements of Davenport and Bain (1935) at a very late stage of ageing, where it seems that only one process is predominant. Thus, the activation energy determinations in this work do not prove that the atomic processes involved are other than the diffusion of carbon and nitrogen.

Another effect may result from this hardening of the individual grains of the matrix. The elongation to fracture of strain-aged specimens is very much less than the virgin material, and decreases the longer the ageing time. This lowering of the ductility is quite unexpected, and cannot be deduced from either the dislocation mechanism, or from the grain-boundary theory alone. Further work on the properties of strain-aged single crystals will be needed to explain this anomaly.

§ 5. THE NATURE OF THE GRAIN BOUNDARY FILM

It is not possible from this work to give any definite answer on the nature of the grain-boundary film, but in the conclusion to this paper it will be convenient to discuss some of its necessary properties.

In the first place, the yield strength of the grain-boundary film must not be too high, for in this case the grains cannot deform at all and the material becomes brittle. Examples are found in copper and gold, alloyed with small amounts of bismuth. Secondly, the grain-boundary film must be continuous. This is the greatest drawback to the old cementite boundary hypothesis for, while some grain-boundary layers of cementite are observed in microsections, they are rare. There is, of course, the possibility that the films are too thin to be observed microscopically, but even here there is the objection that it would take an appreciable time to nucleate new cementite at the grain boundaries during the ageing process. No such induction interval is observed. Thus, the possibilities of a second phase at the grain boundaries must be precluded.

It is therefore necessary to seek a solution to this problem in terms of the carbon and nitrogen which is present, not as cementite or iron nitride, but held in solution in the matrix. Wert (1950) has recently shown, by internal friction methods, that the solubility of carbon in α-iron at room temperature is of the order of 10^{-7} weight per cent. A result of this order is to be expected from Hägg's rule. The ratio of the atomic diameters of iron and carbon is 0·63, whereas the limiting value, according to Hägg, is 0·59. Thus, carbon will exist preferentially in the α-iron lattice at places where the lattice is distorted, and large distorted areas occur at the grain boundaries. It seems possible that carbon will accumulate in these regions, forming areas of carbon-rich material which may have the necessary mechanical properties.

Evidence for grain-boundary segregation is apparently supplied by the etching characteristics of mild steel. If a specimen is so treated by any normal etchant, the grain boundaries come into relief as ridges, which are easily seen if a plastic replica of the surface is taken and examined under the electron microscope. The photographs of Trotter, McClean and Clews (1949) may be taken as typical examples. These ridges do not seem to be cementite, as there is no sharp boundary delineating the second phase. It would be interesting if it could be shown that these ridges are of carbon-rich material, of higher strength than the matrix, for films of this type would give rise to the yield point extension observed in mild steel.

ACKNOWLEDGMENTS

The writer wishes to tender his thanks to Professor Sir Lawrence Bragg, Dr. E. Orowan and Mr. W. M. Lomer for their interest and advice, and to the Royal Commission for the Exhibition of 1851 and the Lord Rutherford Memorial Fellowship for financial support. To Dr. F. C. Frank, of the University of Bristol, the author is deeply indebted for supplying him with the results of the theoretical calculations mentioned in this paper.

REFERENCES

BILBY, B. A., 1950, *Sheet Metal Ind.*, **27,** 707.

COTTRELL, A. H., 1948, *The Strength of Solids* (London : Physical Society), p. 37; 1949, *Progress in Metal Physics* (London : Butterworth's Scientific Publications), p. 105.

COTTRELL, A. H., and BILBY, B. A., 1949, *Proc. Phys. Soc.* A, **52,** 49.

COTTRELL, A. H., and CHURCHMAN, A. T., 1949, *J. Metals* (*Metals Trans.*), **185,** 877.

DAVENPORT, E. S., and BAIN, E. C., 1935, *Trans. Amer. Soc. Metals*, **23,** 1047.

ESHELBY, J. D., FRANK, F. C., and NABARRO, F. R. N., 1951, *Phil. Mag.*, **42,** 351.

FAST, J. D., 1950, *Rev. Métall.*, **47,** 779.

GRIFFITHS, A. A., 1920, *Phil. Trans. Roy. Soc.* A, **221,** 163.

HALL, E. O., 1951, *Proc. Phys. Soc.* B, **64,** 742.

HOLDEN, A. H., and HOLLOMON, J. H., 1949 a, *J. Metals* (*Metals Trans.*), **185,** 179; 1949 b, *Ibid.*, **185,** 878.

KURODA, M., 1938, *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **34,** 1528.

MORRISON, J. L. M., 1939, *Proc. Instn. Mech. Engrs.*, **142,** 193.

NABARRO, F. R. N., 1948, *The Strength of Solids* (London : Physical Society), p. 38.
NADAI, A., 1924, *Z. tech. Phys.*, **5**, 371.
SYLWESTROWICZ, W., and HALL, E. O., 1951, *Proc. Phys. Soc.* B, **64**, 495.
TROTTER, J., MCCLEAN, D., and CLEWS, C. J. B., 1949, *Electron Microscopy Symposium* (London : Institute of Metals), p. 75.
WERT, C. A., 1950, *J. Metals* (*Metals Trans.*), **188**, 1242.
WERT, C. A., and ZENER, C., 1949, *Phys. Rev.*, **76**, 1169.
WINLOCK, J., and LEITER, R. W. E., 1937, *Trans. Amer. Soc. Metals*, **25**, 163.

Reversible Effects in the Magnetization of Nickel

BY R. S. TEBBLE, W. D. CORNER* AND J. E. WOOD
Physics Department, University of Leeds

Communicated by R. Whiddington; *MS. received* 24*th April* 1951

ABSTRACT. An account is given of investigations on the temperature dependence of the contribution from reversible processes to the magnetization of annealed and strained nickel wires. This involves the measurement of reversible susceptibility κ_r over a range of field, and of temperature ($-50°$ C. to $+200°$ C.), and a rapid automatic method of recording (κ_r, H) curves has been developed. The effect of temperature on κ_r is discussed in relation to Becker's strain theory and reasons for the discrepancies are suggested. It is shown that the contribution of reversible processes to the total change in magnetization, integrated over the hysteresis cycle, increases with temperature and strain, from 15% for an annealed specimen at $-50°$ C. to 85% for a strained specimen (17·1 kg. mm^{-2}) at 150° C. The effect is not uniform over the whole cycle and in the region of the coercive field a considerable part of the change in magnetization must be attributed to irreversible processes.

§ 1. INTRODUCTION

In a recent paper (Tebble and Corner 1950, to be referred to as I) an account is given of investigations on the reversible susceptibility κ_r of ferromagnetic materials at room temperature. Estimates were made, *inter alia*, of that part of the change in magnetization which takes place reversibly as the specimen is taken round a hysteresis cycle.

The method involves the measurements of κ_r at various points on the magnetization (I, H) curve. The area under the (κ_r, H) curve is given by

$$\int_{-H_m}^{+H_m} \kappa_r \, dH = \int_{-H_m}^{+H_m} \left(\frac{dI}{dH}\right)_r dH = \Delta I_r, \qquad \ldots\ldots(1)$$

where H_m is the maximum applied field in each direction. This expression gives the ***minimum*** contribution of reversible processes to the total change in magnetization ΔI over the range of field considered. This method of estimating the reversible component of the change in magnetization makes possible the examination of the effect of temperature on the relative importance of the two groups of processes, reversible and irreversible. (The significance of the terms 'reversible' and 'irreversible' is discussed in I, §§ 1 and 2.) For this purpose an automatic method has been developed whereby (κ_r, H) curves can be recorded rapidly while the temperature of the specimen (in the range $-50°$ C. to $+200°$ C.) is held constant and measured accurately. It should be mentioned that a direct experimental examination of the effect of temperature on the irreversible contribution, by measurements on the Barkhausen effect, has been considered, but has been rejected as being impracticable. In order to obtain the required information it is necessary to carry out experiments over several weeks (cf. Tebble, Skidmore and Corner 1950); and it would be extremely difficult to maintain the specimen at the desired temperature for such a period under the conditions of the experiment. If, however, by the method described below, the reversible change in magnetization is correctly estimated the remaining part of the magnetization may be taken as irreversible.

---

* Now at the University of Durham.

This paper is primarly concerned with the presentation of new experimental results and with a limited discussion of their significance. It is hoped that a more comprehensive discussion will be presented later.

§ 2. PREVIOUS WORK

The variation of reversible susceptibility with temperature has been examined by a number of workers, but most of the measurements have been made on specimens in the demagnetized state for which κ_r is usually referred to as the initial susceptibility κ_i. Kirkham (1937) investigated the effect of temperature on the reversible susceptibility of nickel at a number of points on the initial magnetization curve, making use of a ballistic method (see I). It was found that in general κ_r increased with temperature almost up to the Curie point, except for an anomalous effect at about 200° c., ascribed to a change in the easy direction of magnetization of the specimen; near the Curie point κ_r fell rapidly to zero. The work of Kahan (1938) was directed primarily to a study of the validity of the Rayleigh law connecting susceptibility κ with applied field H:

$$\kappa = a + bH, \qquad \ldots\ldots(2)$$

where H is small. The variation of κ_i with temperature showed an anomaly at about 200° c., similar to that observed by Kirkham.

Theoretical treatments of reversible susceptibility have been developed but only from the earlier theories of magnetization, e.g. the 'internal strain' theory of Becker, and Kersten's *Fremdkörpertheorie*, and there is little point in discussing the implications of these in view of Néel's (1946) criticism of the validity of the premises of these treatments. Unfortunately there has as yet been no successful application of Néel's own 'disperse field theory' to the problem of reversible susceptibility. For a discussion of this subject and of the general background to the present work reference may be made to Stoner (1950). A particular case for which the simple strain theory appears to be still valid is that of a material subject to a sufficiently large external stress. Nickel has a negative saturation magnetostriction coefficient λ, and when the anisotropy introduced by an applied tension Z is large compared with the natural anisotropy of the material the easy directions of magnetization lie at right angles to the direction of the tension. On the application of a magnetic field parallel to the direction of the tension the magnetization may change almost entirely by reversible rotations from the easy direction. When the conditions are such that $Z\lambda \gg K$, the anisotropy coefficient, the reversible susceptibility is given by (Becker and Döring 1939)

$$\kappa_r = I_s^2/3Z\lambda, \qquad \ldots\ldots(3)$$

where I_s is the saturation intensity of magnetization.

Becker and Döring (1939) have compared the values of κ_r from this expression with experimental results for nickel under stress, and the agreement, in so far as the variation with tension is concerned, is excellent. The effect of temperature was considered by comparing the values of I_s^2/λ obtained from measurements of I_s and λ at various temperatures with measured values of κ_r. Whereas I_s^2/λ remains practically constant as temperature is increased, there is a marked increase in κ_r, especially near the Curie point. This discrepancy is discussed further below (§§3 and 4).

In none of these investigations has there been a complete examination of the variation with temperature, over a complete hysteresis loop, of both the reversible susceptibility and the intensity of magnetization.

§ 3. EXPERIMENTAL METHOD

The method is based on the fact that the introduction of a ferromagnetic material into a two-coil system increases the mutual inductance. As a result the E.M.F. induced in the secondary by the alternating field of the primary changes, and this change is a measure of the susceptibility of the material. Provided the amplitude of the alternating field is sufficiently small, the reversible susceptibility is given by

$$\kappa_r = \frac{A}{4\pi a}\left(\frac{V_2 - V_1}{V_1}\right), \qquad \ldots\ldots(4)$$

where A and a are the areas of cross section of the secondary coil and specimen respectively, and V_1 and V_2 the E.M.F.s induced in the secondary coil without and with the specimen respectively. The necessary corrections for finite amplitude of alternating field and for eddy current and demagnetizing effects have already been discussed in I.

The circuit is shown in Figure 1. The alternating field is applied parallel to the axis of the cylindrical specimen by means of a solenoidal coil (the primary) through which an alternating current, 5 kc/s., is passed to produce fields of the order of 10^{-4} to 10^{-3} oersted r.m.s. Mounted coaxially within this is the secondary coil, wound on a long glass tube, inside which the specimens in the form of wires are placed, and through which a suitable liquid can be passed at the required temperature. Surrounding the primary and secondary coils is a solenoid (the 'polarizing coil') which provides a steady field, so that the specimen can be brought to any desired part of the magnetization curve.

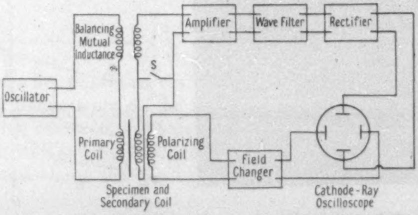


Figure 1. Block diagram of apparatus. Polarizing coil: 3,492 turns. Primary coil: 1,015 turns. Secondary coil: 10,000 turns.

The voltage induced in the secondary coil is amplified and rectified, and the rectified D.C. voltage, which is proportional to V_2, is displayed on the Y plates of a cathode-ray oscilloscope. Included in the primary and secondary coil circuits is a mutual inductance, by means of which the voltage induced without the specimen (V_1) can be balanced off at the input of the amplifier. With the specimen inserted the output voltage from the amplifier is thus directly proportional to $V_2 - V_1$, i.e. to the reversible susceptibility of the specimen. With the switch S closed the secondary coil is shorted, and the output voltage is then proportional to V_1. From these measurements of $V_2 - V_1$ and V_1, and with A and a known, κ_r can be calculated (equation (4)).

The X plates of the oscilloscope are connected to a source of voltage in phase with the current through the polarizing coil, so that the X deflection is proportional to the polarizing field H. This field is changed from a maximum in one direction to a maximum in the other, over a period of about four minutes. Thus as the specimen is taken round a hysteresis cycle, a curve showing the variation of κ_r with applied field is traced out on the oscillograph screen and recorded photographically. The double-beam cathode-ray tube (4 in. diameter) was specially chosen for sharpness of focusing (spot diameter <0·5 mm.) but it was still necessary to apply corrections for non-linearity of the tube and photographic recording; the operation of the rest of the apparatus requires no special comment. The overall accuracy of this method was to within about 1%; a typical uncorrected photograph is shown in Figure 2.

For measurements above room temperature the liquid used as a heating agent was butyl phthallate (B.P. ~350° c.), which was pumped through the specimen tube after electrical heating. Below room temperature use was made of acetone cooled in an acetone–CO_2 mixture.

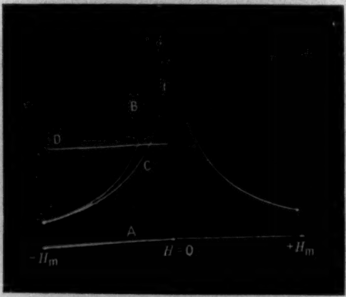


Figure 2. Typical uncorrected (κ_r, H) photograph.
A, base line; D, switch S closed, deflection $\propto V_1$
B, initial (κ_r, H) curve
C, cyclic (κ_r, H) curve } deflection $\propto (V_2 - V_1)$.

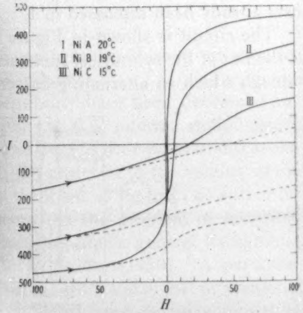


Figure 3. Hysteresis curves for nickel specimens. Full curves show values of I; broken curves show

$$\int_{-H_{\max}}^{H} \kappa_r \, dH.$$

§4. EXPERIMENTAL RESULTS

The results given here are for specimens of nickel wire (diameter 0·046 cm. length 50 cm.) of the following percentage composition, supplied by the Mond Nickel Co.: C 0·04, Si 0·05, Fe 0·15, Mn 0·03, Al 0·01, Ti nil, Mg 0·13, S 0·003, Ni balance. The chemical composition approximates closely to that of the specimen used in I and by Tebble, Skidmore and Corner (1950).

Ni A was annealed *in vacuo* for 24 hours at 1,000° c. then cooled slowly at 120° c. per hour. This specimen was extremely soft mechanically and magnetically unstable, but after a few trial measurements and consequent unintentional cold working it became more stable.

Ni B and C were the same wires, subjected to longitudinal tensions of 3·43 kg. mm^{-2} and 17·1 kg. mm^{-2} respectively, during the measurements (the yield stress of annealed nickel is given in tables (Smithells 1949) as 6 kg. mm^{-2} and the breaking stress as 30 kg. mm^{-2}).

Hysteresis loops were obtained, for each specimen, at a number of temperatures and the results are summarized in Figure 3 and Table 1. These measurements were made under the same conditions as the corresponding susceptibility measurements, i.e. a (κ_r, H) curve was first recorded and, if the temperature was suitable, a hysteresis curve was then obtained at or near the same temperature. Neither the hysteresis loops nor the (κ_r, H) curves have been corrected for the effect of the demagnetizing field of the specimens as this is very small. Four strands of wire 50 cm. long were used in each case and the effective length-to-diameter ratio was 270 : 1, giving a demagnetizing factor of $7{\cdot}6 \times 10^{-4}$ for $\mu = \infty$ (Bozorth and Chapin 1942).

Table 1. Magnetic Characteristics of Specimens at Various Temperatures

	Specimen A ($Z=0$ kg/mm^2)					Specimen B ($Z=3{\cdot}43$ kg/mm^2)				
Temperature (° c.)	−46	20	110	148	202	−49	19	98	138	184
I ($H=100$ oe.)	435	467	452	431	391	357	360	352	346	335
I (remanence)	226	223	219	194	179	183	198	159	147	151
H_c	1·60	1·23	1·07	0·94	0·90	6·6	6·5	5·6	5·4	4·8

	Specimen C ($Z=17{\cdot}1$ kg/mm^2)			
Temperature (° c.)	−51	15	81	161
I ($H=100$ oe.)	162	164	162	158
I (remanence)	65	36	36	30
H_c	24·6	19·6	17·6	12·6

§ 5. RESULTS

(i) (κ_r, T) *Curves*

Room temperature (κ_r, H) curves for two specimens are given in Figure 4, and the effect of temperature on the form of the curves is shown for specimen Ni A. For clarity, curves at only three of the ten temperatures are given. The (κ_r, H) curves for Ni B and Ni C change similarly with temperature, as is indicated in Figure 5, which shows the effect of temperature on κ_r at maximum (i.e. at $H \simeq H_c$).

Even for maximum strain there is still a marked variation of κ_r with temperature, which, on the Becker treatment, is inconsistent with the experimental result that I_s^2/λ is practically constant (equation (3)). In view of the magnitude of the applied tension in nickel C, it would seem rather unlikely that internal forces could play any appreciable part in the magnetization of this specimen. This point is discussed further in the following section but it is fairly certain that at high values of applied fields (of the order of 100 oersted) the magnetization of nickel C takes place by reversible rotations of the magnetization vector as is assumed in equation (3). Nevertheless, the temperature dependence of κ_r is as great in this region as at low field values. It should be pointed out, however, that, although considering specimens under strain, Becker and Döring have taken values of λ which are those for an unstrained specimen and it is quite possible that this is the cause of part of the discrepancy. In view of the unquestioned validity of the arguments leading to equation (3), it would be of interest to have more information on the effect of tension on the temperature dependence of λ.

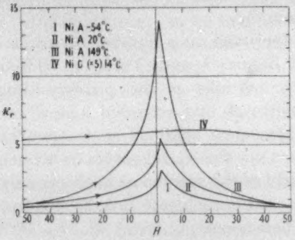


Figure 4. Variation of reversible susceptibility κ_r with polarizing field H.

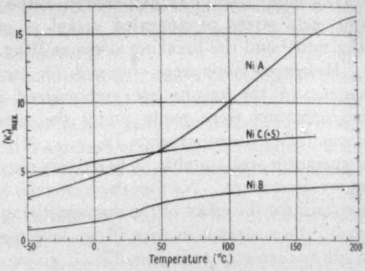


Figure 5. Effect of temperature on maximum reversible susceptibility $(\kappa_r)_{max}$.

(ii) *Reversible Changes in Magnetization*

Reversible susceptibility is in itself a measure of the contribution of reversible processes to the change in magnetization for a small increment in field. It can be seen from a (κ_r, H) curve that the reversible contribution is a maximum near the region $H = H_c$. The relative importance of reversible processes, however, must be assessed by comparison with the total change in magnetization taking place, as indicated by the hysteresis curve. Two methods of showing this have been adopted.

In the first, as illustrated by Figures 6 to 8, the variation is shown of κ_r/κ with field, where $\kappa = dI/dH$, and I is the intensity of magnetization as given by the hysteresis loop. Since $\kappa_r = dI_r/dH$, the Figures 6 to 8 show the relative importance of reversible processes (as a fraction of the total) at different parts of the magnetization curve. (In order to avoid confusion the results of only two temperatures are given in Figure 8, as between room temperature and 160° c. the curves overlap.)

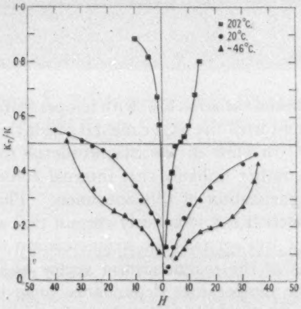


Figure 6. Variation of κ_r/κ with polarizing field—NiA.

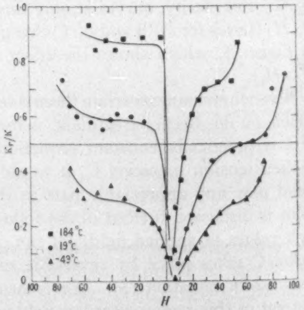


Figure 7. Variation of κ_r/κ with polarizing field—NiB

In the second method, the (κ_r, H) curves are integrated as indicated in equation (1) over the range of field considered ($H_m = 100$ oersted), and values of ΔI_r obtained over a range of temperature for each specimen. Since the total change in magnetization ΔI as measured on the hysteresis loop also changes with temperature, the results have been shown in Figure 9, with $\Delta I_r/\Delta I$ plotted against temperature. ΔI_r has also been estimated for intermediate values of polarizing field H by evaluating the integral $\int_{-100}^{H} \kappa_r \, dH$ and the results are shown in Figure 3 (broken curves).

The effect of increased temperature on an annealed specimen, such as Ni A, will be to reduce the anisotropy and local internal strains, and thus increase the range of reversibility of the microscopic boundary movements making up the changes in magnetization. This is borne out by the results shown in Figure 9 where $\Delta I_r/\Delta I$ increases with temperature. However, the effect is not uniform over the hysteresis curve as can be seen in Figure 6, from which it would appear that in the region $H = H_c$ there is a large part of the change in magnetization which is not reversible. Unless this method of measurement results in a considerable under-estimate of reversible effects, the difference $\Delta I - \Delta I_r$ must be attributed to irreversible processes, which are present even at temperatures of 200° c. At high values of H, however, (~100 oersted) it appears that most of the change in magnetization can be considered reversible (Figures 3, 6, 7, 8).

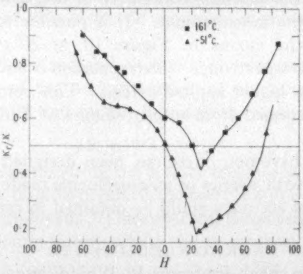


Figure 8. Variation of κ_r/κ with polarizing field—NiC.

In a specimen such as Ni C, with a stress of 17·1 kg.mm^{-2}, it has usually been tacitly assumed that the magnetization takes place over the whole magnetization curve mainly by reversible rotations of the magnetization vector from the directions of easy magnetization (as in Becker and Döring's treatment). Such would certainly be the impression gained from the shape of the hysteresis curve. In the region of the coercive field, at least, this would appear to be an over-simplification since some non-reversible processes are obviously taking place (Figure 8), presumably because of the effect of internal strains and impurities. The occurrence of discontinuous effects has been confirmed by the results of experiments which show that there is an appreciable Barkhausen effect in Ni C.

(iii) (κ_r, I) *Curves*

In Figure 10 the reversible susceptibility κ_r is shown as a function of the intensity of magnetization I, taken from the corresponding hysteresis loop. This method has certain advantages as a means of showing the variation of susceptibility over the hysteresis loop in that both variables are functions of the properties of the material itself. These curves serve as a useful check on the accuracy of both the hysteresis loops and (κ_r, H) curves, particularly near the coercive field where even small errors would obviously result in noticeable irregularity in the (κ_r, I) curves. What is possibly one of the most important properties of such curves is that extrapolation to the value $\kappa_r = 0$ should give a value of I equal to I_s the saturation intensity of magnetization of the material at the temperature of the measurement. It is possible to extrapolate with fair precision for two of the curves in Figure 10, at 20° c. and 149° c., to give $I_s = 486 \pm 4$ and 436 ± 4 respectively. (Extrapolation of the $-50°$ c. curves would require measurements at higher applied fields.) This compares with the values $I_s = 485$ and 435, determined from measurements of I in very high fields (see Stoner 1936).

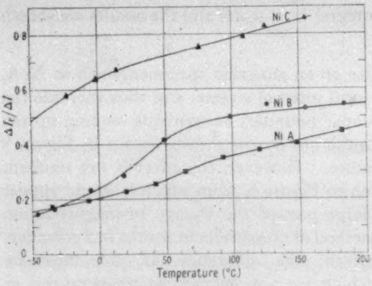


Figure 9. Effect of temperature on reversible contribution $\Delta I_r/\Delta I$.

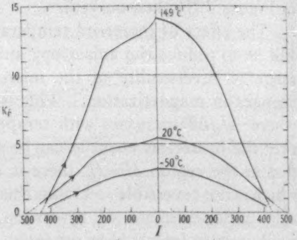


Figure 10. (κ_r, I) curves for annealed nickel (NiA) - - - - - extrapolated.

These experiments have not, of course, been designed for this purpose, and it is very probable that with a series of measurements made with that end in view, a considerable degree of accuracy could be obtained in estimating I_s.

ACKNOWLEDGMENTS

The authors wish to thank Professor R. Whiddington, under whose helpful direction this work has been carried out, Professor E. C. Stoner and Dr. P. Rhodes for their generous advice in the preparation of the paper, and the Mond Nickel Co. Ltd., who supplied the specimens. One of us (J.E.W.) is indebted to the Department of Scientific and Industrial Research for a maintenance grant.

REFERENCES

Becker, R., and Döring, W., 1939, *Ferromagnetismus* (Berlin : Springer), p. 130.
Bozorth, R. M., and Chapin, D. M., 1942, *J. Appl. Phys.* **13,** 320.
Kahan, T., 1938, *Ann. Phys., Paris,* **9,** 105.
Kirkham, D., 1937, *Phys. Rev.,* **52,** 1162.
Néel, L., 1946, *Ann. Univ. Grenoble,* **22,** 299.
Smithells, C. J., 1949, *Metals Reference Book* (London : Butterworth's Scientific Publications), p. 554.
Stoner, E. C., 1936, *Phil. Trans. Roy. Soc.* A, **235,** 165; 1950, *Rep. Prog. Phys.,* **13,** 138 (London : Physical Society).
Tebble, R. S., and Corner, W. D., 1950, *Proc. Phys. Soc.* B, **63,** 1005.
Tebble, R. S., Skidmore, I. C., and Corner, W. D., 1950, *Proc. Phys. Soc.* A, **63,** 739.

The Surface Tension of Supercooled Phenyl Ether

By C. DODD
University College, London

Communicated by E. N. da C. Andrade; *MS. received* 11*th May* 1951

ABSTRACT. Determinations of the surface tension of phenyl ether have been made, using Jaeger's method, for the liquid in both the ordinary and the supercooled state.

No abnormal change with temperature, of the type already found for viscosity and dielectric constant, has been detected for surface tension as the liquid enters the supercooled region.

§1. INTRODUCTION

Viscosity measurements on phenyl ether (Dodd and Hu 1949) indicate the existence of a structural change occurring at the melting point; the points obtained by plotting $\log\eta$ against $1/T$, where η is the viscosity at absolute temperature T, fall into two groups corresponding to observations made above and below the melting point, the points of each group lying on a straight line but the line for the supercooled liquid being steeper than that for the liquid above the melting point. A similar effect has also been observed in the temperature variation of the dielectric constant of this substance (Dodd and Roberts 1950).

In order to see if this structural change at the melting point extends to the surface, measurements of the surface tension of phenyl ether have been made both for the normal liquid and for the liquid in the supercooled state.

§2. APPARATUS

The method used for measuring the surface tensions at different temperatures was Jaeger's maximum bubble-pressure method; Figure 1 shows diagrammatically the arrangement employed. The apparatus is so designed that the specimen liquid is not in direct communication with the atmosphere, so that contamination of the surface is avoided and conditions are ideal for supercooling. The jet C, consisting of a short length of precision bore capillary tube, is sealed into the glass vessel G containing the specimen liquid L, and the reservoir D also contains a little of this liquid in order to prevent loss of liquid from G by evaporation.

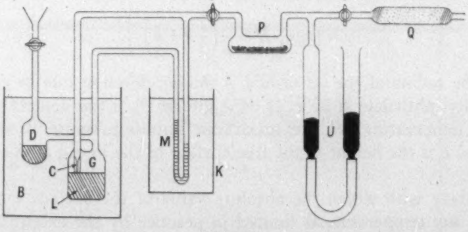


Figure 1.

Both G and D are immersed in a large, glass-sided water bath B whose temperature can be maintained constant to within 0·01° c. at any temperature within the range 0° c. to 50° c. The manometer M containing butyl phthalate —a liquid with a very low vapour pressure at 20° c.—is immersed in the tank K having optical glass sides and containing water maintained at a temperature of $20{\cdot}00 \pm 0{\cdot}02$° c. The valve V regulates the supply of air to the jet, this supply being dried by passage through the drying tubes P and Q containing phosphorus pentoxide and calcium chloride respectively. The tubes U contain mercury and the right-hand limb is free to move in a vertical plane.

§ 3. METHOD

The apparatus G is first tilted so that the plane of the jet is horizontal—a state of affairs which can be judged accurately by eye when the jet is just below the surface of the liquid L. With the water bath B at a temperature of 20° c. just enough phenyl ether is added to the vessel G for the plane of the jet to coincide with the free surface of the liquid in that vessel. The space above the mercury in the left-hand limb of the U tube is filled with clean, dry air drawn in through the drying tube Q by raising and then lowering the right-hand limb with V closed and the tap T open. Fixing this limb so that the mercury levels are as in the diagram, the valve V is opened so that bubbles issue from the jet into the liquid L at the rate of about one every two minutes. The maximum difference of levels in the manometer M is measured by keeping the cross-wires of a cathetometer reading to 0·001 cm. on the slowly moving meniscus of butyl phthalate until the bubble is released and repeating the observation ten times. This difference of levels gives the excess pressure inside the bubble above that in the space above L and is independent of the actual pressure inside the apparatus.

This procedure is carried out with the bath B at various temperatures both above and below the melting point of the phenyl ether. It was possible to continue this procedure even with the liquid supercooled more than twenty degrees below its melting point. At temperatures below 20° c. the plane of the jet is above the free surface of the liquid L, due to thermal contraction; nevertheless the effect of surface tension maintains contact between the jet and the liquid throughout.

The value for the surface tension of the liquid is calculated by means of the equation

$$\gamma = \tfrac{1}{2} a g \rho \left[H - \frac{\sigma}{\rho}\left(\frac{2}{3}a + h\right)\right],$$

where a is the radius of the jet orifice, g the acceleration due to gravity, ρ the density of butyl phthalate at 20° c. (1·0458 gm.cm^{-3}), σ the density of the liquid at the bath temperature, H the maximum bubble-pressure in cm. of butyl phthalate, and h is the height of the free surface of the liquid above the plane of the jet orifice.

The accuracy with which the absolute value of the surface tension can be calculated at any temperature is limited in practice by the extent to which the section of the jet orifice is circular and by the accuracy with which the mean radius of this section can be measured. With such a small jet (radius 0·0102 cm.) this accuracy is not great, being of the order of one part in two hundred.

However writing the equation in the form $\gamma = A[H - \Delta H] = A[H']$ it is evident that since A is an apparatus constant, changes in surface tension due to temperature changes are proportional to changes in the length H'. In these experiments,

H is of the order of 5 cm. whilst $\Delta H = (\sigma/\rho)(\frac{2}{3}a + h)$ is only 0·065 cm. at the maximum temperature used, so that a value for H' can be obtained to an accuracy of 1 part in 5,000 assuming only an approximate value for the radius of the jet in order to calculate ΔH. Similarly the value of h, which is by definition zero at 20°c., is easily shown to be given by the product of the depth of the liquid in G, the temperature excess over 20°c., and the coefficient of apparent expansion of the liquid relative to Pyrex. For this apparatus the change in h is only 0·00191 cm. for one degree change of temperature and need only be estimated to an accuracy of one part in two hundred.

§ 4. RESULTS

The values of H' at temperature θ°c. both above and below the melting point are given in the Table, and the variation of surface tension with temperature is shown in Figure 2 where H' is plotted as a function of the temperature. It is

Table

θ (°c.)	H (cm.)	H'_{obs} (cm.)	$H'_{\mathrm{obs}} - H'_{\mathrm{calc}}$ ($\times 10^{-3}$ cm.)	θ (°c.)	H (cm.)	H'_{obs} (cm.)	$H'_{\mathrm{obs}} - H'_{\mathrm{calc}}$ ($\times 10^{-3}$ cm.)
49·70	4·803	4·738	0	26·85	Melting Point		
47·75	4·831	4·770	+2	25·80	5·120	5·102	+2
46·00	4·854	4·797	+4	24·30	5·137	5·122	−1
44·80	4·870	4·816	+3	23·35	5·150	5·137	0
44·00	4·878	4·825	0	22·00	5·169	5·158	+1
43·10	4·891	4·840	+2	19·70	5·196	5·190	−3
42·00	4·903	4·854	0	18·65	5·210	5·206	−3
40·30	4·925	4·879	−1	17·00	5·232	5·231	−3
38·15	4·951	4·909	−3	15·60	5·258	5·259	+4
36·50	4·975	4·936	−1	12·90	5·292	5·299	+2
35·35	4·990	4·953	−2	12·20	5·298	5·306	−1
34·00	5·010	4·975	−1	10·50	5·321	5·322	−1
32·40	5·030	4·998	−2	8·70	5·342	5·357	−4
31·30	5·045	5·015	−1	8·30	5·350	5·366	−1
30·05	5·062	5·035	0	5·90	5·382	5·403	−1
28·20	5·084	5·060	−2				
27·10	5·099	5·078	+2				

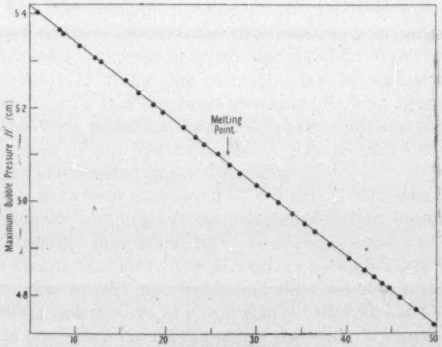


Figure 2.

seen that all points lie on a single line and there is no indication of any discontinuity in slope occurring at the melting point. This is emphasized in the Table, where the errors obtained by fitting the values of H' to the equation $H' = 5{\cdot}494 - 1{\cdot}536 \times 10^2\theta + 3{\cdot}2 \times 10^{-6}\theta^2$ are also given. Over the whole range these errors are small and unsystematic, showing that there is no abnormal change in surface tension as the liquid enters the supercooled region. This is also brought out in Figure 3 where surface tension over (density)$^{2/3}$ plotted

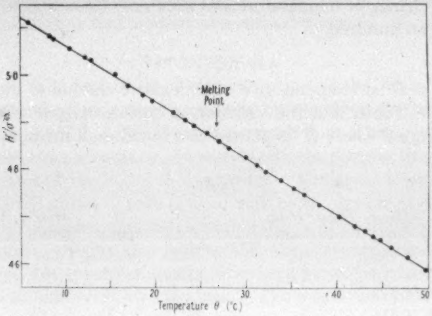


Figure 3.

against temperature gives a single *straight* line over the whole range showing the applicability of the Eötvös formula $\gamma/\sigma^{2/3} = K(t - t_c)$ where t_c is the critical temperature.

Using the above equation for H' and the mean value for the jet radius obtained using a travelling microscope, the surface tension of phenyl ether at its melting point (26·85°c.) is 26·6 dyne cm^{-1} with values of 28·7 and 24·8 dyne cm^{-1} at 0°c. and 50°c. respectively.

REFERENCES

Dodd, C., and Hu Pak Mi, 1949, *Proc. Phys. Soc.* B, **62**, 454.
Dodd, C., and Roberts, G. N., 1950, *Proc. Phys. Soc.* B, **63**, 814.

Freezing of Supercooled Water

By A. W. BREWER and H. P. PALMER
Clarendon Laboratory, Oxford

MS. received 19*th March* 1951

ABSTRACT. Rau has claimed that, by repeated freezing and thawing, the temperature to which a drop of water supercools before freezing can be lowered and, with special care, the freezing temperature can be reduced to −72° c. This is claimed to be the temperature of spontaneous nucleation of ice. At that temperature the ice forms in rectangular crystals which also melt at −72° c. This note describes a careful attempt to repeat Rau's work. Like Cwilong, we have been unable to do so except by deliberate contamination, and it is presumed that Rau's apparatus was faulty and his drops were contaminated by his cooling fluid. Information is given concerning the reproducibility of the freezing temperature of individual drops when repeatedly frozen and thawed. Rau's rectangular crystals have been observed in strong alcohol solutions but no information concerning their nature has been obtained.

---

§ 1. INTRODUCTION

THERE is now a relatively extensive literature on the freezing of supercooled water, and a partial bibliography is appended to this paper. Martin (1913) supercooled water to −26° c., and Mayer and Pfaff (1935) to −33° c., Cwilong (1947) to −23° c., Smith-Johannsen (1948) to −38° c., but the most striking result which has been described is that of Rau (1944), who claimed that by repeated freezing and thawing the temperature at which a drop of water would freeze could be lowered. He claimed, further, that under special conditions water could be supercooled to −72° c., and at this temperature cubic ice crystals appeared which actually melted if warmed above −72° c.

Rau claimed that the incidence of freezing in supercooled water is due to the presence of freezing nuclei each of which acts at its own characteristic temperature, and that although freezing nuclei can cause freezing repeatedly, they gradually lose their effectiveness after remaining in water or damp air, especially if the water is repeatedly frozen and thawed. Rau supercooled drops of water to about −25° c. by this technique, but he reports also that drops remained liquid to −72° c. if they were condensed from air which had stagnated in his apparatus for a sufficient time, and whose "nucleus content had been rendered innocuous by suitable measures". Rau tabulated his results in ten points which are summarized in the Appendix.

These results have been questioned by Cwilong (1947), who was only able to reproduce the results with ether or acetone contamination, but whose paper has, unfortunately, received little attention.

In private communications German workers have criticized Cwilong's work on the ground that he did not freeze and thaw his drops sufficiently often. We have therefore endeavoured to repeat Rau's work and, like Cwilong, we have been unable to reproduce his results unless the water is contaminated. Rau's results can be reproduced with ease in water contaminated with alcohol.

It must be particularly noted that in this paper the freezing temperature is the temperature at which freezing begins. On account of the difficulties of the nucleation of the ice phase this temperature is usually well below 0° c., which, of course, is the *melting* temperature of pure ice.

§ 2. APPARATUS

Rau made his observations on water drops formed by condensation from what was claimed to be dust-free air, on polished metal plates each about 3 cm. × 4 cm., which were in the upper part of a temperature-controlled chamber of approximately five litres capacity, the drops being viewed through a window in the top of the chamber. The air in the chamber was stirred by a fan driven by an electric motor which was also in the chamber. The body of the chamber was cooled by 'a coolant' presumably alcohol, or some similar organic solvent, which was circulated round the chamber and through a heat exchanger cooled by solid CO_2. The metal plates were cooled from the same heat exchanger but were independently controlled. Rau states that for the lower temperatures liquid-air cooling was employed, but he does not say whether, in this case, the other coolant was entirely eliminated. The apparatus is to be criticized because, first, contamination from the circulating fluid can occur, and second, the fan motor, unless brushless, will produce nuclei from arcing and wear at the commutator; it will also stir up any dust in the chamber.

For our work, to observe the variation of the freezing temperature on repeated freezing and thawing, the apparatus was particularly designed to avoid these faults. It was also made fully automatic, to freeze and thaw a drop in a continuously repeated cycle, and to record the temperature of freezing on each cycle. This was done to avoid any criticism that the drop had not been frozen a sufficient number of times.

The apparatus is shown diagrammatically in Figure 1. The drop under test was one particular, selected, drop in a water deposit on the gold surface S. The surface was backed by the copper 'mushroom' M and its temperature could be controlled by balancing a variable electrical heating against the cooling due to conduction through the brass stem, and heavy copper rod C. The copper rod dipped into liquid oxygen in the Dewar vessel. The electric heater was wound on the upper part of the cooling stem at H. The actual surface used was formed by soldering a thin sheet of the material selected on to the top of the copper mushroom and the temperature was measured by a thermocouple between this sheet and the copper.

The surface was contained in an airtight enclosure. The gas pressure in the enclosure was always maintained at a value slightly above atmospheric and the difference measured by a mercury manometer. This was to prevent any 'dirty' air leaking into the space and providing nuclei which could settle on the surface. (The freezing temperature of the drops rises if the surface is exposed to the atmosphere for a short time.) As far as practicable dead space in the enclosure was filled with brass spacing washers W. This was because when the temperature of the surface is raised water evaporates from the drops, since the enclosure tends to contain air saturated at the surface temperature, and at the high temperatures more water is required for this. Thus as the temperature of the surface is varied through a cycle the drops tend also to vary through a cycle in size, and this effect is reduced by reducing the dead space. Rau stated that freezing nuclei can be reactivated by drying out, and we therefore wished to reduce the risk of drying to a minimum. The enclosure could be ventilated by passing air through the inlet and outlet pipes shown. In the inlet system there was an arrangement of traps and filters by which the air could be humidified and filtered. Most of the work was carried out in oxygen obtained by evaporating liquid oxygen. This was largely because it is a convenient source of a gas, at pressures slightly above atmospheric, which is absolutely free from condensable vapours.

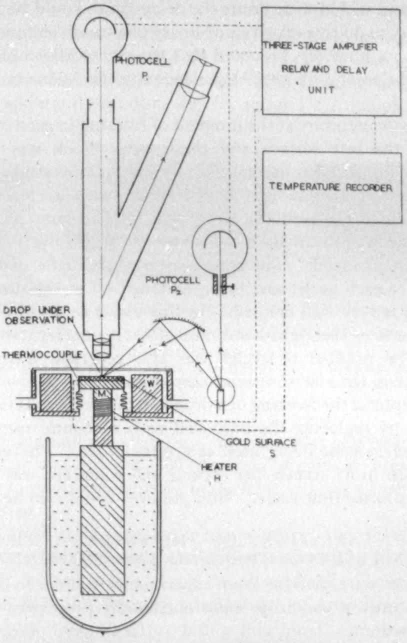


Figure 1. Diagram of apparatus

The drop of water on the surface S whose properties were being examined was observed through the low-power binocular microscope. A photocell P_1 was placed in one eyepiece and the object could be examined visually through the other. The drop was very brilliantly illuminated by condensing upon it the image of the arc of a high-pressure mercury lamp. If the surface is clean and well polished and the drop is clear liquid, the field of view is relatively dark. If the drop is considerably supercooled, then, on freezing, it becomes brilliant white very rapidly indeed as dendritic crystals spread through the drop. Water spaces are left between the dendrites and are filled in later as the drop loses sufficient heat for all the water to freeze. At $-20°$ c. a $\frac{1}{2}$ mm. drop undergoes its initial whitening, corresponding to much of the total brightness change in less than 0·02 second. The effect of this brightness change on the photocell was amplified and used to operate 'telephone' relays in such a way that, after freezing, the temperature of the drop was held roughly steady for about 20 seconds. This was to ensure that the whole drop froze, after which the heater was turned on fully till the drop melted. The melting reduced the brightness again, affecting the photocell, and it was arranged that the temperature of the surface was again held steady for about 20 seconds, in order to ensure that all the drop had melted. The heater was then switched off and the surface cooled to freeze the drop and so repeat the cycle. All rates of heating and cooling could be fully controlled, and within wide limits the delay times could be regulated easily. The arrangements to do this involved ordinary electronic techniques and are not described. It may, however, be noted that the photocell–amplifier system was balanced, using the second photocell to balance any effects due to lamp-brightness fluctuations.

To record the temperature at the moment of freezing in each cycle the thermocouple between the test surface and the copper block was connected to a galvanometer; an ice bath was used as the reference temperature. The lamp and galvanometer were in a light-tight box together with a clock drum carrying bromide paper which replaced the usual scale. The lamp was covered by a stop with a pinhole in it, the image of which moved across the photographic paper as the temperature changed. The lamp was arranged to be switched on during the cooling part of each cycle, and to be switched off at the instant of freezing, which, of course, is very well defined. In this way a series of parallel lines was recorded on the trace, the right-hand end of each line representing a freezing temperature. The number of cycles was also recorded independently by a telephone counter in the electronic circuitry.

Ciné photographs of the freezing of a drop and of the crystals in a contaminated drop were taken by replacing the photocell with a 16 mm. camera, the second eyepiece of the microscope being used as a view-finder. Photographs could be taken at speeds up to 64 frames per second and a marker was available in the camera to give a precise time scale. Still pictures could also be taken.

§ 3. RESULTS: VARIATIONS OF THE FREEZING TEMPERATURE ON REPEATED FREEZING AND THAWING

Several surfaces were used for these experiments in order to establish that the freezing temperatures of the drops were not merely properties of one metal or of a polishing medium. Gold and nickel surfaces were used, and the nickel was polished in turn with rouge and water, diamantine and water, and metal polish. The freezing temperatures of the drops did not appear to be affected by these changes of surface. After polishing, all surfaces were washed with freshly condensed steam. The surfaces obtained in this way were all bright and clear and represent the best approach we could make to the surfaces used by Rau. They were all more or less greasy as the water did not wet the surface but formed drops upon it.

(i) '*Clean*' *Conditions—No Airborne Nuclei*

The surface was carefully cleaned and all the apparatus ventilated thoroughly with oxygen, free from all condensation nuclei. The oxygen was obtained by evaporating commercial liquid oxygen, moistening it by passing it over distilled water, and filtering with a cotton-wool filter. The cotton wool was ordinary absorbent cotton wool, and auxiliary experiments with a small cloud chamber similar to an Aitken nucleus counter showed that the filtered oxygen was free from all condensation nuclei.

The surface was then cooled and water accumulated on the test surface either as dew or, by lowering the surface temperature below about −20° c., the water could be accumulated as hoar frost, which was subsequently melted to form drops on the surface. In either case the surface was covered with drops of a wide range of sizes, and usually the process was stopped when the largest drop had grown to about 1 mm. diameter. The apparatus would only work automatically on drops larger than about 0·3 mm. diameter, but there was usually a good selection of these.

When grown under clean conditions drops larger than about 0·2 mm. diameter froze in a regular way at a temperature near to −20° c. *and up to* 500 *cycles of freezing and thawing had no effect whatever*; the freezing temperature usually fluctuated in an apparently random way by about 1° c., though sometimes these fluctuations were as small as ± 0·25° c. The first part of a typical trace in which a drop was frozen and thawed 500 times is shown in Figure 2. *We were absolutely unable to obtain any lowering of the freezing point whatever by the repeated freezing and melting.*

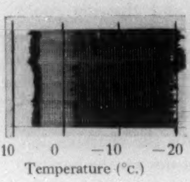


Figure 2. The variation of the temperature of freezing of a drop of pure water frozen repeatedly.

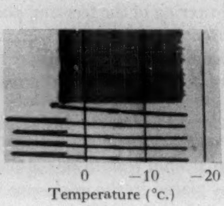


Figure 4. The effect of changing the rate of cooling of a drop. There is no change in the freezing temperature.

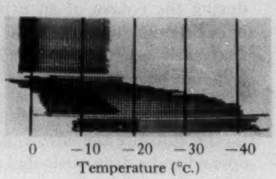


Figure 5. The effect of introducing alcohol into the chamber. The temperature of freezing is lowered as the alcohol distils into the drop.

The smaller drops freeze at a lower temperature; the smallest drop in which freezing can be detected by visual examination is about 30 μ (0·03 mm.) diameter, and the freezing temperature of such drops is usually about −30° c. There was no evidence that the freezing temperature of any drops, large or small, could be affected by the cycling process.

Under supercooled conditions the drops, once freezing has started, freeze very rapidly indeed. Figure 3 (Plate *) shows two successive frames of a ciné film taken at 64 frames/second. It will be seen that the drop has turned white between the first and second frames as the crystals have grown through the drop in less than 0·02 second. This is to be compared with Rau's illustrations. His figures 2 and 6 show apparently stationary crystals in supercooled water; in pure supercooled water we have been quite unable to reproduce this.

(ii) *Dependence on the Rate of Freezing*

For truly isolated drops there is no effect on changing the period of a cycle over a ratio of more than 10 to 1. The longest period we have used is about 15 minutes and the shortest 1 minute. Figure 4 shows a trace in which the period of the cycle was changed, but it will be seen that the freezing temperature remains at −17° c. throughout.

* For Plates see end of issue.

If the drop is not truly isolated, but is close to a drop which for some reason is freezing at a higher temperature, then the freezing temperature can appear to become erratic, because very fine, hair-like, dendritic crystals grow out of the nearby drop which has frozen, towards the water drop, and eventually touch it; the water drop then freezes immediately. Thus if there is one drop on the surface which is frozen, then by simply holding the surface at any temperature below 0° c. every drop on the surface can be frozen by the growth of dendrites from the ice into the water drops. Apart from this the rate of cooling had no effect whatever on the freezing temperature.

§4. THE EFFECTS OF CONTAMINATION WITH ALCOHOL

It is well known that the addition of any solute lowers the melting temperature of ice, and it is obvious that the lowering of the freezing point which Rau observed during the course of an experiment on repeated freezing and thawing could have been due to the gradual accumulation of coolant in a leaky apparatus. Rau frequently asserted that this is impossible, though the reasons for this statement have not been given.

(i) *Alcohol Contamination without Airborne Nuclei*

To simulate the effects of an alcohol leak in our apparatus the chamber was opened and a drop of alcohol was placed in the enclosure but not on the surface. When the surface was cooled the alcohol distilled on to it and contaminated the water drops. Alcohol–water drops free from atmospheric nuclei still supercooled about 20° c. below their melting point, but as the melting point dropped with gradually increasing amounts of alcohol, so did the freezing point (cf. Hollstein (1947), who obtained the same result with common salt solutions). Figure 5 shows the effects obtained and in particular the lowering of the freezing and melting temperatures, both moving together. This is to be compared with curves 1 and 2 of Rau's figure 3, which show how he found the freezing temperatures were lowered. The similarity will be recognized, but it will be noted that Rau's drops froze initially at $-2°$ c. and $-5°$ c., which suggests the presence of very active nuclei preventing significant supercooling.

(ii) *The Effect of Airborne Nuclei without Alcohol*

If the lid of the enclosure is removed for a while and atmospheric dust and nuclei are allowed to settle on the surface, then the freezing temperature of many drops will rise markedly, to a temperature as high as $-5°$ c. Drops which contain freezing nuclei of this kind usually prove, during continuous cycles of freezing and thawing, to be erratic in their freezing temperature, which will sometimes fall as low as about $-20°$ c. This is the usual freezing temperature of a drop when not contaminated with airborne nuclei. Sometimes the freezing temperature will stay at about $-20°$ c. for a long time, but sometimes it will recover to a much higher freezing temperature. The changes are usually in large jumps rather than in steps as described by Rau. *By repeated freezing and thawing we have never lowered the freezing temperature to a value below that which would have been obtained directly by excluding airborne nuclei.*

(iii) *Reproduction of Rau's Results*

Curves 1 and 2 of Rau's figure 3 refer to drops which froze initially at about $-3°$ c., but whose freezing temperatures fell to nearly $-20°$ c. when the drops had been frozen and thawed fifteen times. It is probable that these drops were condensed on to the surface from air freshly introduced into the apparatus, and this air must have contained some very active nuclei. Any coolant which had leaked into the apparatus would accumulate on the surface only when, in the course of an experiment, it was cooled below the temperature of the cooling bath. Drops contaminated with active nuclei and then with increasing amounts of coolant are necessary to reproduce these results precisely.

Rau attained very much lower freezing temperatures when the drops were condensed from air which had "stagnated for a sufficient length of time" in the apparatus, and comparatively large amounts of coolant could leak into the apparatus during that time. The cubic crystals appearing at −72° c. could be reproduced in solutions of about 70% alcohol, which contained many airborne nuclei.

§ 5. CRYSTAL PHENOMENA

As was noted above, the water drops in our apparatus always froze very rapidly, at about −20° c. if clean. The drop became brilliantly white in less than 0·02 second as dendritic crystals spread right through it. It is quite impossible to photograph the crystal phenomena, as Rau claims to have done. Smith-Johannsen (1948) has reported that in his apparatus samples of water usually froze rapidly and became white, but occasionally single columnar crystals formed slowly which were only visible under polarized light. No explanation can be offered for this effect, for our drops *always* froze very rapidly, and it is difficult to see how any ice crystal can grow slowly in pure supercooled water.

Drops contaminated with both active nuclei, whereby significant supercooling does not occur, and alcohol, which slows down and stabilizes the crystal growth, show stationary crystal phenomena which can be photographed with ease. The phenomena follow those described by Rau and are detailed below.

(*a*) *0°c. to −20°c.: weak alcohol solutions.* Crystals are in the form of needles, frequently as rosettes. The photographs shown in Rau's figure 2 are very typical.

(*b*) *−20°c. to −50°c.* In stronger alcohol concentrations hexagonal plates are formed with hexagonal pyramid ends exactly as shown in photographs 1 and 2 of Rau's figure 6. Figure 6 (Plate) shows a photograph which we have obtained. The ends are actually *hollow* pyramids, and when the plates melt they first become rings as the thinner centres disappear first. At the lower temperatures crystal growth is very slow and, even in the presence of nuclei, the solutions readily supercool on that account.

(*c*) *−72°c.* In the stronger solutions (about 70% alcohol), at a temperature near −72° c., but depending somewhat upon the concentration, a new crystal structure appears. Rectangular plates are formed with square pyramid ends. The crystal growth is very rapid and, unless the solution is cooled slowly, a mush of fine crystals is obtained. The crystals melt, as Rau describes, at about −72° c., according to their temperature of formation. Figure 7 (Plate) shows a rectangular crystal. As in Rau's figure 7, the rectangular pyramid ends can be seen. The behaviour is exactly as Rau describes, and hexagonal plates can usually be grown from the same solution at about −50° c. The rate of growth of the hexagonal crystals is very slow indeed at −70° c. The difference in the rate of growth of the cubic and hexagonal crystals is very striking, and it is difficult to see how two crystal forms of pure ice could grow at such different rates at the same temperature.

The solutions in which the rectangular crystals appear correspond approximately to equal molecular concentrations of alcohol and water, and the crystals could easily be alcohol monohydrate $C_2H_5CH.H_2O$. Attempts to isolate the crystals and analyse them have been unsuccessful, as the mother liquid is virtually a glass, and hexagonal water crystals grow during the filtering.

Some workers, who used x-ray and electron diffraction techniques, suggest that ice may have a cubic structure below −70° c., but others report no change in the structure from 0° c. to −183° c. (see bibliography given by Lonsdale and Owston 1948). Both Rau and Cwilong obtained crystals by sublimation below −72° c., and these crystals appeared to be cubic; Cwilong's were formed from pure water vapour, and Rau's therefore may have been, but we have been obliged to conclude that the cubic crystals which Rau formed in a liquid cannot be formed in pure liquid water.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the interest which the Reader in Meteorology at Oxford University, Professor G. M. B. Dobson, has shown in this work, and we are indebted to the Professor of Experimental Philosophy, Lord Cherwell, for enabling us to carry out this work in the Clarendon Laboratory. One of us (H.P.P.) is under grant from the Meteorological Office, and we are indebted to the Director, Sir Nelson Johnson, for permission to publish these results.

APPENDIX

Rau's conclusions	*Comments*
1, 2, 3. The incidence of freezing in supercooled water is due to the presence of freezing nuclei each of which acts at its own characteristic temperature. There is a whole spectrum of such nuclei.	Agreed.
4, 5. A freezing nucleus can cause freezing repeatedly, but gradually loses its effectiveness after remaining in water or damp air. It regains its activity when dried.	A freezing nucleus in pure water remains fully effective when it remains in water for 24 hours, and the water is frozen 500 times in that period.
6, 7, 8. If freezing nuclei are excluded, water can be supercooled to −72° c., at which temperature cubic crystals form. These melt when warmed above −72° c.	This cannot be reproduced in pure water, but all these phenomena appear in strong alcohol solutions containing many freezing nuclei.
9. A marked diminution of the surface tension of water occurs below −55° c., indicating a change in the state of association.	This probably indicates the condensation of alcohol or acetone on to the surface below this temperature.
10. Below −72° c. cubic crystals can be formed by sublimation from water vapour.	No comments. See bibliography given by Lonsdale and Owston (1948).

REFERENCES

Bangham, D. H., 1946, *Nature, Lond.*, **157,** 733.

Cwilong, B. M., 1947, *Proc. Roy. Soc.* A, **190,** 137, *J. Glaciol.*, **1,** 53.

Dorsey, N. E., 1948, *Trans. Amer. Phil. Soc.*, **38,** 247.

Frank, F. C., 1946, *Nature, Lond.*, **157,** 267.

Heverley, J. R., 1949, *Trans. Amer. Geophys. Un.*, **30,** 205.
Hollstein, E., 1947, Referred to by H. Weickmann, *The Ice Phase in the Atmosphere*, Royal Aircraft Establishment Translation No. 273, p. 20.
Lonsdale, K., and Owston, P., 1948, *J. Glaciol.*, **1,** 118.
Martin, W. H., 1913, *Trans. Roy. Soc. Can.*, **111,** 219.
Meyer, J., and Pfaff, W., 1935, *Z. anorg. Chem.*, **224,** 305.
Rau, W., 1944, *Schr. Deut. Akad. Luft.*, **8,** 65.
Schulz, G., 1948, *Met. Rdsch.*, **1,** 237.
Smith-Johannsen, R., 1948, *Science*, **108,** 652.
Ubbelohde, A. F., 1946, *Nature, Lond.*, **157,** 625.

Spontaneous Condensation of Water Vapour in Expansion Chamber Experiments

By B. J. MASON
Department of Meteorology, Imperial College, London

Communicated by Sir David Brunt; MS. received 25th January 1951

ABSTRACT. When clean, saturated air is subjected to a sufficiently large adiabatic expansion, small water droplets arise spontaneously from the vapour. An equation is derived for the growth rate of a droplet in an atmosphere of given supersaturation and temperature. This is used in conjunction with the Becker–Döring theory of nucleation to calculate the supersaturation and the concentration of droplets at successive time intervals during a cloud-chamber expansion, allowance being made for warming of the air by liberation of the latent heat of condensation. For large expansion ratios, the supersaturation attains a maximum and the droplet concentration a sensibly constant value *before* the end of the expansion. The computed value of this maximum droplet concentration is in acceptable agreement with the recent observations of Frey. The experimental results appear incompatible with Tolman's recent theory concerning the variation of surface tension with droplet radius. The predicted rates of droplet growth agree well with recent measurements made by Hazen.

§ 1. INTRODUCTION

It was found by C. T. R. Wilson (1897) that when air initially saturated with water vapour and freed as far as possible from impurities and ions was subjected to a sufficiently large adiabatic expansion, a cloud of very small water droplets appeared spontaneously from the vapour. A cloud, as distinct from a few scattered droplets, appeared only if the expansion ratio exceeded 1·37, producing momentarily an eightfold supersaturation. Since then, many workers have sought to redetermine this 'critical' supersaturation at which liquid aggregates arise from the initially homogeneous vapour, obtaining inconsistent results. The discrepancies may perhaps be attributed to (*a*) differences in the degree to which truly adiabatic expansions were achieved, leading to different estimates of the final supersaturation, (*b*) variation in the degree of purity of the air and vapour, and, perhaps more important, (*c*) the absence of a common criterion for the onset of condensation. Indeed, the adoption of such a criterion is not easy since it appears that no sharp transition should, nor does in fact, occur.

In the absence of foreign nuclei, aggregates of the condensed phase will be formed by chance collisions of molecules of the supersaturated vapour; if these aggregates attain a critical size they will survive and continue to grow—otherwise they will disappear. The critical radius r_c of a spherical aggregate is given by

$$r_c = (2\sigma M/\rho \boldsymbol{R}T)\ \ln\,(p/p_\infty), \qquad \ldots\ldots(1)$$

where σ is the surface tension and ρ the density of the condensed phase, M its molecular weight, $\boldsymbol{R}$ the gas constant, T the absolute temperature, p the pressure of the supersaturated vapour and p_∞ the equilibrium vapour pressure at temperature T over a plane surface of the liquid.

The methods of statistical thermodynamics allow calculation of the probability of the formation of aggregates of critical size, and hence the rate of production of nuclei for development of the liquid phase. The relevant expression has been derived by Becker and Döring (1935) viz.:

$$J = \frac{ZXv}{r_c}\left(\frac{3W_c}{\pi \boldsymbol{k}T}\right)^{1/2} \exp\left(-\frac{W_c}{\boldsymbol{k}T}\right), \qquad \ldots\ldots(2)$$

where J is the rate of formation of nuclei/cm^3 sec., Z the number of vapour molecules/cm^3, X the number of molecules striking 1 cm^2/sec. multiplied by the condensation coefficient, v the volume occupied by a molecule in the liquid, $W_c(=4\pi r_c^2\sigma/3)$ the work of nucleus formation and $\boldsymbol{k}$ Boltzmann's constant. For easier comparison with experiment, (2) may be written

$$\ln J = \ln\frac{\alpha}{\boldsymbol{R}^2\rho}\left(\frac{2\boldsymbol{N}^3 M\sigma}{\pi}\right)^{1/2} + 2\ln\frac{p_\infty}{T} + 2\ln\frac{p}{p_\infty} - \frac{16\pi M^2\sigma^3}{3\rho \boldsymbol{k}\boldsymbol{R}^2 T^3 \ln^2 p/p_\infty} \qquad \ldots\ldots(3)$$

in which α is the condensation coefficient and $\boldsymbol{N}$ is Avogadro's number.

It is seen from (3) that J should increase continuously and rapidly with the saturation ratio p/p_∞, as shown in curve *a* of Figure 1, plotted for expansions commencing at $T = 290°$ K. Volmer and Flood (1934) checked equation (3) by determining experimentally the value of p/p_∞ at which the rate of nucleation J attained a value of 1/cm^3 sec., and obtained very good agreement between theory and observation. With $T = 261°$ K. the experimental and calculated values of the critical saturation ratio were 5·03 and 4·96 respectively; the nucleation rate was observed to increase rapidly at higher supersaturations.

A continuous rise of droplet concentration with increasing supersaturation was observed earlier by Andrén (1917) who employed expansion ratios of between 1·253 and 1·544 to produce droplet concentrations of between 1/cm^3 and 10^5/cm^3; he observed a rapid increase in the number of droplets when the expansion ratio exceeded 1·37.

§ 2. THE EXPERIMENTS OF FREY

Recently, Frey (1941) has made a careful determination of the total concentration of droplets as a function of the *nominal* * maximum supersaturation. The expansion chamber was filled with samples of air which, having been freed from dust and ions and allowed to attain room temperature, were expanded to various final volumes. The duration of an expansion was about 0·01 second. The droplets produced in a well-defined and brightly illuminated volume of the chamber were photographed through a microscope. Time-lapse photographs were taken at 64 frames per second for a period of 1 second after the expansion.

* See § 4.

The minimum concentration which could be accurately measured was $10^3/\text{cm}^3$, corresponding to about one droplet in every ten pictures. Curve *b* of Figure 1 shows the number of droplets appearing in unit volume as a function of the nominal saturation ratio p/p_∞. If the curve is extrapolated as shown one can deduce that a concentration of $1/\text{cm}^3$ would occur with $p/p_\infty = 4{\cdot}9$; such a saturation ratio would be achieved at a temperature of 263° K. in Frey's experiments. Comparison of this result with the prediction of equation (3) involves knowing the time for which the supersaturation persists after cessation of the expansion. Here, Frey gives no information, but agreement between his results and the theory would require this period to be 1 second; the dimensions of his chamber suggest however, that its sensitive time could not exceed 0·1 second, so that Frey's measured value of the critical saturation ratio is rather lower than the corresponding theoretical value. With increasing nominal supersaturation, Frey found the droplet concentration increased continuously—rapidly at first, then more slowly—until finally it became constant at $p/p_\infty \sim 14$. The maximum concentration of droplets was about $2 \times 10^6/\text{cm}^3$ (see Figure 1, curve *b*).

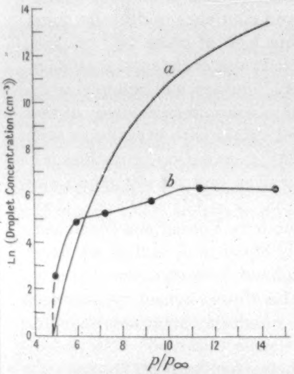


Figure 1. Curve *a*. Rate of nucleation as a function of the saturation ratio. Curve *b*. Measured concentration of droplets as a function of the nominal saturation ratio (*after* Frey).

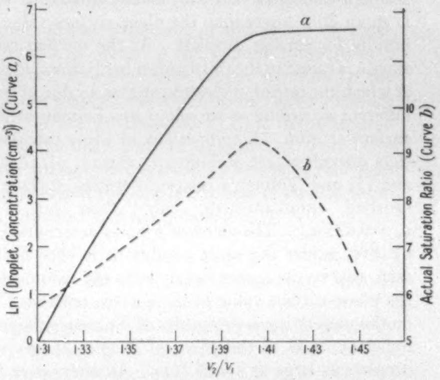


Figure 2. Calculated values of the droplet concentration and supersaturation as a function of the expansion ratio. Curve *a*. Droplet concentration. Curve *b*. Saturation ratio.

The main purpose of this paper is to provide a theoretical explanation of Frey's results. First, however, his own interpretation must be discussed.

§ 3. EFFECT OF DROPLET RADIUS ON SURFACE TENSION

Frey seeks to explain his observations of no further increase of droplet concentration after the supersaturation attains a certain high value in terms of an earlier theory by Lenard (1914) that, for small droplets, the surface tension will have a value different from that for a plane surface of the same liquid. On this basis, Frey argues that for each saturation ratio greater than 5·4 (the critical value for the complex $(H_2O)_2$), there are two characteristic nuclear sizes, but finally a supersaturation is reached when *all* the existing embryos grow to visible drops. Frey claims that this is the case at $p/p_\infty = 14{\cdot}5$, in agreement with the experimental value at which he observed the droplet concentration to become constant. Frey's arguments are not at all clear, but they appear to depend on the exact form of a curve connecting droplet sizes and corresponding values of their surface tensions, and on the ratio of the actual droplet size at a given supersaturation to that defined by equation (1). Quantitative evaluation of these relationships are apparently based on some early measurements of ionic mobilities by Przibram (1908).

A classical theory of the effect of droplet size on surface tension has recently been worked out by Tolman (1949), his results being confirmed by Kirkwood and Buff (1949) and Koenig (1950) using statistical-mechanical methods. Tolman calculates that the surface tension of water droplets of 10^{-7} cm. radius is about 20% lower than the plane-surface value and continues to decrease quite rapidly for smaller droplets. As the nuclear radius r_c is of order 10^{-7} cm. and since σ is raised to the third power in (3), the calculated value of the supersaturation at which the rate of nucleation attains a value of, say, 1/cm³ sec. will be appreciably different according as we adopt the Tolman or the plane-surface value σ_∞ of the surface tension. A comparison of these calculated values with experiment will thus provide a test of Tolman's theory. If $T = 261°$ K. and $J = 1$, equations (1) and (3) and Tolman's relation between surface tension and droplet radius are satisfied simultaneously only when $p/p_\infty = 3{\cdot}55$, $r_c = 8{\cdot}35 \times 10^{-8}$ cm. and $\sigma_{r_c} = 0{\cdot}825\ \sigma_\infty$. The value of p/p_∞ as determined both by Volmer and Flood and by Frey under the same conditions is very nearly equal to 5, and as we have seen, this result agrees closely with the value calculated from equation (3) using the plane-surface value of the surface tension. This throws considerable doubt on the underlying assumptions of Tolman's theory which appears to overestimate the magnitude of the effect of droplet radius on surface tension—at least for droplets as large as $r = 10^{-7}$ cm. An alternative theory for the observed rates of droplet formation will now be given, in which the plane-surface value of the surface tension is assumed throughout.

§ 4. TREND OF THE SUPERSATURATION DURING EXPANSION

In plotting curve *b*, Figure 1, Frey calculated the saturation ratio $S(=p/p_\infty)$ from the formula

$$S = \frac{p_1}{p_2}\frac{v_1}{v_2}\frac{T_2}{T_1}, \qquad \ldots\ldots(4)$$

where p_1, v_1 and T_1 are respectively the equilibrium vapour pressure, the volume and temperature at the beginning of the expansion, the suffix 2 denoting corresponding values on its termination. A knowledge of v_1, v_2 and T_1 allows the other three quantities to be calculated if the expansion is assumed to be adiabatic. It is generally assumed that S so calculated corresponds to the maximum supersaturation attained during the expansion; for given initial conditions it will increase with the expansion ratio v_2/v_1 and substitution in equation (3) predicts higher concentrations of droplets with larger expansion ratios. It must be remembered, however, that during an expansion the concentration of vapour does not remain constant since it is continually removed by condensation; also the air is warmed by the latent heat liberated during condensation. Hence, the supersaturation will reach its peak before the end of a large expansion and thereafter continue to fall as more and more vapour becomes involved in droplet formation. If condensation occurs, the maximum supersaturation will be less than the *nominal* supersaturation calculated from equation (4), although the difference will be inappreciable until large numbers of droplets are formed (see Figure 2).

§ 5. GROWTH OF THE DROPLETS

The growth rate of a droplet will be determined by the rate of transfer of water vapour to and the latent heat of condensation from its surface. The rate of increase of droplet radius r as controlled by diffusion of water vapour may be expressed in the form $r\,dr/dt = (DM/\rho\boldsymbol{R}T)\{p - \bar{p}(r)\}$ or

$$\frac{p - p_s(T_r)}{p_s(T)} = \frac{\rho \boldsymbol{R} T}{p_s(T) D M}\, r \frac{dr}{dt}, \qquad \ldots\ldots(5)$$

where ρ is the density of the liquid, D the coefficient of diffusion[†] of water vapour in air, M the molecular weight of water, p the vapour pressure at a considerable distance from the droplet, $\bar{p}(r)$ the mean vapour pressure in the immediate vicinity of the droplet (assumed equal to the equilibrium vapour pressure $p_s(T_r)$ at the surface temperature T_r of the droplet) and $p_s(T)$ is the saturation vapour pressure at temperature T in the main body of vapour.

While the droplet is growing the surface temperature T_r will be higher than that of the surrounding air. If it is assumed that the latent heat of condensation is dispersed solely by conduction through the air, we have

$$r\,dr/dt = (K/L\rho)(T_r - T), \qquad \ldots\ldots(6)$$

where K is the thermal conductivity of the air[†], and L the latent heat of condensation.

The variation of saturation vapour pressure with temperature is given by $(1/p_s)dp_s/dT = JLM/\boldsymbol{R}T^2$ which integrated from the droplet surface to the main bulk of the vapour gives

$$\ln \frac{p_s(T_r)}{p_s(T)} = \frac{JLM}{\boldsymbol{R}} \frac{(T_r - T)}{TT_r} \simeq \frac{JLM(T_r - T)}{\boldsymbol{R}T^2}. \qquad \ldots\ldots(7)$$

Thus from (6),

$$\frac{p_s(T_r) - p_s(T)}{p_s(T)} = \exp\left(\frac{JL^2 M\rho}{K\boldsymbol{R}T^2}\, r \frac{dr}{dt}\right) - 1. \qquad \ldots\ldots(7a)$$

Adding (5) and (7*a*),

$$\frac{p - p_s(T)}{p_s(T)} = \exp\left(\frac{JL^2 M\rho}{K\boldsymbol{R}T^2}\, r \frac{dr}{dt}\right) + \frac{\rho \boldsymbol{R} T}{p_s(T) D M}\, r \frac{dr}{dt} - 1 \qquad \ldots\ldots(8)$$

or

$$S = \frac{p}{p_s(T)} = \exp\left(\frac{JL^2 M\rho}{K\boldsymbol{R}T^2}\, r \frac{dr}{dt}\right) + \frac{\rho \boldsymbol{R} T}{p_s(T) D M}\, r \frac{dr}{dt}. \qquad \ldots\ldots(9)$$

[†] As the dimensions of the droplet are of the same order as the mean free path of the air molecules, the ordinary values of the diffusion coefficient and thermal conductivity have to be replaced by compensated coefficients which are numerically smaller (Langmuir 1944).

$$D = \left[\frac{r}{D^*(r - \alpha\lambda_{\mathrm{w}})} + \frac{1}{r}\left(\frac{2\pi M}{\boldsymbol{R}T}\right)^{1/2}\right]^{-1}, \qquad K = \left[\frac{r}{K^*(r + \alpha\lambda_{\mathrm{a}})} + \frac{4}{r\rho_{\mathrm{a}} c_{\mathrm{p}} f}\left(\frac{3\boldsymbol{R}T}{M}\right)^{1/2}\right]^{-1},$$

where D^* and K^* are the ordinary coefficients, α is a constant ($=0{\cdot}7$), λ_{w} the mean free path of water molecules, ρ_{a} the air density, c_{p} the specific heat of air at constant pressure, f an accommodation factor of air molecules in collision with liquid water ($f = 0{\cdot}7$) and the other parameters are as defined above. $D \simeq D^*$ and $K \simeq K^*$ when $r > 5 \times 10^{-4}$ cm.

Now if the exponent in (8) is considerably less than unity, we may write as a first approximation

$$\left(\frac{JL^2M\rho}{\boldsymbol{KR}T^2}+\frac{\rho\boldsymbol{R}T}{p_s(T)DM}\right)r\frac{dr}{dt}=S-1. \qquad \ldots\ldots(9a)$$

Equation (9 *a*) can be solved for $r\,dr/dt$, the value of which is substituted in (9) as the first step in solving this equation by successive approximations.

We may compare the growth rate predicted by (9) and (9 *a*) with the measurements of Hazen (1942) who found r^2 to vary linearly with time as predicted by (9 *a*). With ethyl alcohol subjected to a 15% expansion in nitrogen Hazen's measured value of dr^2/dt was $5\times10^{-6}\,\mathrm{cm^2\,sec^{-1}}$; assuming the initial temperature of the vapour to be 290° K., equation (9 *a*) predicts a value of $5\times10^{-6}\,\mathrm{cm^2\,sec^{-1}}$ while (9) gives $4{\cdot}5\times10^{-6}\,\mathrm{cm^2\,sec^{-1}}$.

§ 6. RATE OF DROPLET FORMATION DURING AN EXPANSION

The rate of formation of new droplets at any instant during an expansion will be determined by the prevailing temperature and supersaturation. These latter quantities can be calculated from a knowledge of the initial conditions, the expansion ratio and the concentration and size of existing droplets. Since the rate of droplet formation and the rate of change of supersaturation are interdependent, calculation of the total concentration of droplets formed during the expansion period involves a step-by-step computation. The total number of water molecules will, of course, remain constant; this imposes one of the conditions which must be satisfied at each step of the calculation.

During the expansion the volume is assumed to increase linearly with time; the expansion period is then divided into equal intervals of time corresponding to equal increments of volume. The vapour concentration existing at the end of each interval is calculated by subtracting from the initial concentration the equivalent concentration of the existing droplets, the volume of which is calculated with the aid of (9). The temperature at successive intervals is calculated assuming the expansion to be adiabatic, a correction then being made for the warming of the air by the latent heat of condensation. This allows the mean supersaturation existing at the end of each interval to be computed, substitution of which in (3) allows the number of new droplets appearing during the next interval to be computed. In carrying out this last step it is assumed that the vapour is homogeneous whereas in reality each existing droplet is the centre of a region of lower supersaturation. The assumption of a mean supersaturation existing between these regions and the main bulk of the gas is therefore an approximation. However, when the removal of vapour and the release of latent heat by existing droplets begins to have an appreciable effect on the supersaturation their mean separation is only about $100\,\mu$; calculation shows that 50% of any instantaneous rise in temperature or fall in vapour concentration occurring at the surface of a droplet will appear at a point midway between two droplets in 10^{-4} sec., 80% of the change appearing in 10^{-3} sec. Hence, inhomogeneities in the temperature and diffusion field are largely smoothed out over periods of order 10^{-3} sec. and a mean supersaturation may be assumed in the following calculations without serious error.

The results of one calculation are represented by Figure 2. Air of initial temperature 290° K. is assumed to expand to 1·45 times its original volume in 0·01 sec. The full curve shows the concentration of droplets at successive intervals during this expansion as a function of the expansion ratio. It is seen that the concentration rises sharply as the expansion proceeds, but more slowly when the ratio exceeds about 1·39 and finally becomes sensibly constant at $v_2/v_1=1{\cdot}42$, which corresponds to a nominal saturation ratio of 10·4. The maximum concentration of droplets is $3{\cdot}8\times10^6/\text{cm}^3$ as compared with Frey's value of $1{\cdot}8\times10^6/\text{cm}^3$; the agreement is considered to be good, especially as Frey does not measure accurately the expansion time, nor state the temperature of his samples before expansion which would allow a closer comparison to be made. Moreover, the presence of turbulence in his chamber must have introduced errors in counting the droplets. Also, in carrying out the computations one has to assume values for the surface tension of supercooled water and that the expansion is truly adiabatic. The trend of the supersaturation during the course of the experiment is shown by the dotted curve; for small expansions it is identical to the nominal supersaturation, but when the expansion exceeds 39% the saturation ratio begins to fall below its nominal value and continues to decrease rapidly as condensation progresses.

It appears possible then, to predict theoretically the total concentration of droplets that will be formed during a large expansion, where according to Volmer (1939) the period of droplet formation may be taken equal to the expansion period. With small expansions (small numbers of droplets) however, the supersaturation may persist for a period after the expansion and as Frey does not measure this period, no comparison between the theory and the overall form of his experimental curve can be made.

REFERENCES

Andrén, L., 1917, *Ann. Phys., Lpz.*, **52,** 1.

Becker, R., and Döring, W., 1935, *Ann. Phys., Lpz.*, **24,** 719.

Frey, F., 1941, *Z. phys. Chem.* B, **49,** 83.

Hazen, W. E., 1942, *Rev. Sci. Instrum.*, **13,** 247.

Kirkwood, J. G., and Buff, R. P., 1949, *J. Chem. Phys.*, **17,** 338.

Koenig, F. O., 1950, *J. Chem. Phys.*, **18,** 449.

Langmuir, I., 1944, G.E.C. Report on *Supercooled Water Droplets in Rising Currents of Cold Saturated Air.*

Lenard, P., 1914, *Ber. Heidelberger Akad. Wiss.*, **5,** A 27.

Przibram, K., 1908, *Ber. Akad. Wiss. Wien,* **117,** 678.

Tolman, R. C., 1949, *J. Chem. Phys.*, **17,** 333.

Volmer, M., 1939, *Kinetik der Phasenbildung* (Dresden and Leipzig : Steinkopff), p. 132.

Volmer, M., and Flood, H., 1934, *Z. phys. Chem.* A, **170,** 273.

Wilson, C. T. R., 1897, *Phil. Trans. Roy. Soc.* A, **189,** 265.

Light Flux received by a Spectrograph from a Spatially Extended Refracting Source

By L. A. WOODWARD and J. H. B. GEORGE
Inorganic and Physical Chemistry Laboratories, Oxford

Communicated by R. F. Barrow; *MS. received 14th February* 1951

ABSTRACT. The geometrical-optical treatment of Nielsen, which referred to a spatially extended source of refractive index unity, is extended to sources of any refractive index. Provided that the direction of emergence of the light from the source is nearly normal, it is shown that a change of refractive index from 1 to n modifies the light flux by the factor $1/n^2$. This result is independent of whether or not a condensing lens is used between source and spectrograph. Its relevance in connection with observed intensities in Raman spectra is discussed.

The question of the amount of light received by a spectrograph from a spatially extended source is of interest in relation to various kinds of experiment, including, for example, the measurement of intensities in Raman spectra (Woodward and George 1951). It has been considered by Nielsen (1930) as a problem in geometrical optics, his treatment referring to a uniformly luminous source of refractive index equal to that of air, which may be taken as practically unity. The purpose of the present paper is to extend Nielsen's treatment to sources of any refractive index. Rays coming from such a source are refracted on passing out into the air, with the consequence that the volume of the source 'seen' by the spectrograph is different from that which would be seen for a source with refractive index effectively the same as that of air.

Nielsen shows that the flux ϕ_1 from a source of unit refractive index, bounded at the front and rear by planes normal to the optical axis, is given by

$$\phi_1 = KI_1l_1, \qquad \ldots\ldots(1)$$

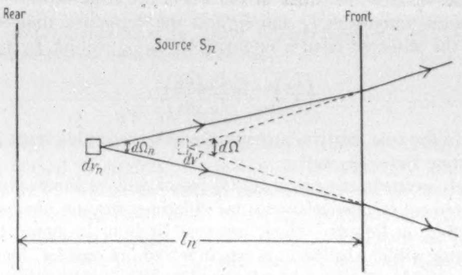
in which K is a function of the dimensions of the collimator and so is a constant for any particular choice, I_1 is a measure of the luminosity of the source in the sense that the amount of light energy emitted in unit time from any small volume element dv_1 in any small element of solid angle $d\Omega_1$ is equal to $I_1\,dv_1\,d\Omega_1$, and l_1 is the depth of the source measured along the optical axis. The other extensions of the source do not enter into the expression for the flux; they must be large enough not to exclude any region that could be 'seen' by the spectrograph. This proviso will be assumed throughout. It is to be noted that the value of the flux, as given above, is then independent of the position of the collimator relative to the source. It is also unaffected by the introduction of a condensing lens between source and collimator, being in particular independent of the focal length and position of such a lens. The only proviso here is that the condensing lens must have a diameter sufficiently large to deal with all rays from the source that could contribute to the flux through the collimator. This will be assumed in what follows.

Consider now the flux ϕ_n from a source S_n of refractive index n (not equal to unity), whose luminosity is I_n and whose depth is l_n. The element of flux from any volume element dv_n in the element of solid angle $d\Omega_n$ is given by

$$d\phi_n = I_n\, dv_n\, d\Omega_n, \qquad \ldots\ldots(2)$$

the total flux ϕ_n being obtained by integrating over the range of Ω_n appropriate to each volume element, and then integrating over all volume elements of the source.

The value of ϕ_n can be obtained by considering what we may call "the equivalent source S′ of unit refractive index". The volume element dv' of S′ corresponding to the volume element dv_n of S_n is the virtual image of dv_n seen by an eye situated in the air in front of the source, and is obtained (see Figure) by producing backwards without refraction the rays from dv_n which have emerged from the source. Assuming that the angle between any such ray and the normal is small, it follows at once from Snell's law that linear dimensions of dv' at right angles to the optical axis are the same as those of dv_n, but that the linear dimension of dv' along the optical axis is $1/n$ times the corresponding dimension of dv_n. Hence



$$dv' = dv_n/n. \qquad \ldots\ldots(3)$$

Also if $d\Omega'$ is the element of solid angle corresponding to $d\Omega_n{}'$ we have

$$d\Omega' = n^2\, d\Omega_n. \qquad \ldots\ldots(4)$$

Since the element of flux is the same for both,

$$I'\, dv'\, d\Omega' = I_n\, dv_n\, d\Omega_n, \qquad \ldots\ldots(5)$$

where I' is the luminosity of the equivalent source S′ corresponding to I_n for the real source S_n. Inserting (3) and (4) into (5), we obtain

$$I' = I_n/n. \qquad \ldots\ldots(6)$$

The depth l' of the equivalent source is given by

$$l' = l_n/n. \qquad \ldots\ldots(7)$$

Hence by Nielsen's result, equation (1), we have for the flux ϕ' from the equivalent source of unit refractive index

$$\phi' = KI'l' = KI_n l_n/n^2. \qquad \ldots\ldots(8)$$

But $K_n I_n l_n$ is the value of the flux ϕ_1 which would have been received by our collimator from a source of refractive index unity having the same luminosity and depth as the actual source of refractive index n. This conclusion may be expressed

$$\phi_n = \phi_1 / n^2. \qquad \ldots\ldots(9)$$

Since, as Nielsen has pointed out, the flux from a source of refractive index unity is independent of the position of the collimator and is unaffected by the introduction of a condensing lens, it follows that the same must be true for a source of any refractive index. The one condition for the validity of the result expressed in equation (9) is that the light must emerge nearly normally from the source. This condition is well satisfied, for example, in all ordinary Raman effect work.

The result obtained is of interst in connection with the experimental determination of the relative intensities of sources of the same depth. For two such sources with the intensities I_A and I_B and the refractive indices n_A and n_B, respectively, the observed relative intensity $(I_A/I_B)_{obs}$ will be, by equation (9),

$$\left(\frac{I_A}{I_B}\right)_{obs} = \frac{I_A}{I_B}\left(\frac{n_B}{n_A}\right)^2.$$

Thus to obtain the true relative intensity the observed value must be multiplied by the correction factor $(n_A/n_B)^2$.

This purely geometrical-optical (G-O) factor will be involved, for example, in the measurement of relative intensities of light-scattering phenomena such as the Raman effect in liquids. Here, however, it is to be noted that the G-O correction factor will not be the only one involved. Consider, for example, the intensities of scattering of a particular Raman line by a particular molecular species at the same volume concentration in two different solvents, the solutions A and B having different refractive indices. Suppose that the intensity of the exciting source is the same for both solutions and that there are no complications due to chemical interaction with the solvent, association and the like. Then, because of the G-O factor, the observed intensities will not be expected to be equal. But, even when this factor has been properly taken into account, it is not to be expected that the true intensities I_A and I_B will be equal, because of different effects of the media upon the actual excitation of the scattering effect. As is well known, the intensity of Raman scattering is proportional to the square of the amplitude of the exciting electric field strength. Although for both solutions the exciting light is of the same intensity (i.e. has the same field strength) *in air*, the field strengths actually acting upon a molecule in the different media will be different, and so will give rise to different intensities of Raman scattering. (It seems likely that such effects of the medium will act in the opposite sense from the G-O correction.)

For liquids it is at present impossible to calculate reliably the magnitude of these internal field effects. Due application of the G-O correction to experimentally determined relative Raman intensities would appear to open up the possibility of obtaining new knowledge about the liquid state. Experimental investigations along these lines are in progress.

REFERENCES

Nielsen, J. R., 1930, *J. Opt. Soc. Amer.*, **20**, 701.
Woodward, L. A., and George, J. H. B., 1951, *Nature, Lond.*, **167**, 193.

Optical Properties of Selenium

By J. J. DOWD

Research Laboratory, Associated Electrical Industries Ltd., Aldermaston, Berks.

Communicated by E. Billig; *MS. received 7th March* 1951

ABSTRACT. Measurements have been made of the refractive index of amorphous selenium in the red and near infra-red spectral region. The absorption coefficient of amorphous selenium has been measured in the ultra-violet, visible and near infra-red spectral regions. The transmission coefficients of several single crystals of selenium have been measured and an estimate made of the absorption coefficient in the wavelength range 0·68 to 2·0 μ. The results show that the absorption edge of the crystalline material occurs at about the same wavelength as that for the amorphous form. A discrepancy between the value of the near infra-red refractive index and the dielectric constant of amorphous selenium indicates a further absorption band in the infra-red.

---

§ 1. INTRODUCTION

DESPITE the considerable volume of work which has been carried out on the optical properties of selenium in the past fifty years there still seemed to be insufficient optical data available to permit an accurate estimate of either the height and width of the optical absorption band, or about the width of the forbidden energy gap. The object of this work was to determine these data and to correlate the results with those of previous workers as far as possible.

§ 2. PREVIOUS WORK

(i) *Amorphous Selenium*

(*a*) R. W. Wood (1902) measured the absorption coefficient and refractive index of amorphous selenium in the visible spectral region. The absorption measurements were made on films of selenium obtained by cathodic sputtering in high vacuum. In the determination of the refractive index Wood used three thin prisms of selenium in the wavelength region 0·76 to 0·60μ and sputtered films in the region 0·60 to 0·40μ.

(*b*) Meier (1910) measured the refractive index and extinction coefficient of bulk amorphous selenium in the wavelength region 0·257 to 0·668μ by a polarimetric method.

(*c*) Merwin and Larsen (1912) obtained a value of 2·92 for the refractive index at a wavelength of 0·589μ and a value of 2·716 at about 0·67μ. The results were obtained by measurements on selenium prisms.

(*d*) Becker and Schaper (1944) measured the absorption coefficient of amorphous selenium films in the spectral region 0·70 to 0·58μ. The films were obtained by cathodic sputtering.

(ii) *Crystalline Selenium*

There appear to have been no transmission measurements on crystalline selenium. Miller (1925), Skinner (1917) and Weld (1922) have determined the extinction coefficient and refractive index by measurement of the intensities of reflected polarized light. These measurements were made throughout the visible and near ultra-violet spectral region.

§ 3. PRESENT WORK

(i) *Refractive Index of Amorphous Selenium*

The refractive index of amorphous selenium was measured in the spectral region 2·9 to 0·6μ using six prisms of selenium. The refracting angle of these prisms varied from 13° 45′ to 29° 32′.

The prisms were made by pouring molten selenium into a wedge-shaped mould which had two quartz plates as sides. The whole mould was rapidly cooled to reduce crystallization and the prism was then removed from the mould. Examination of the light from a tungsten filament lamp transmitted through the prism enabled one to determine the degree of crystallization which had taken place. The six prisms used in this work were selected for their sparsity of crystallization. The spectroscopically pure selenium was supplied by Johnson Matthey and Co.

Radiation from a Nernst filament was made monochromatic by means of a Hilger Barfit infra-red spectrometer and this monochromatic radiation was focused by means of a convex lens on to the collimator slit of a spectrometer. On the spectrometer was mounted each selenium prism in turn and the angles of minimum deviation were measured for wavelengths between 2·9 and 0·64μ.

The position of minimum deviation in the red region of the spectrum was found by eye. The angles of the prisms were measured by reflection from the two refracting surfaces.

The detector used in the infra-red was a lead sulphide cell. The incident radiation was chopped at a frequency of 820 c/s. and the A.C. signal from the lead sulphide cell was amplified by a homodyne amplifier.* Figure 1 shows a block diagram of the amplifier.

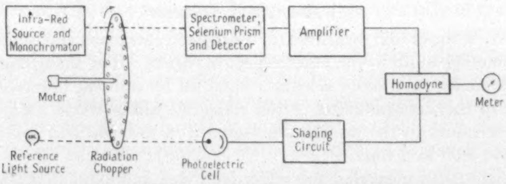


Figure 1. Block diagram of detecting system.

Figure 2 shows the results obtained for refractive index plotted against wavelength, the vertical lines representing the fiducial limits for 95% level of significance.† Where fewer than six results were obtained at a given wavelength no fiducial limits have been plotted. The total experimental spread varied from 0·6 to 2·0%. This spread was mainly due to random experimental errors and not to differences between the specimens. The magnitude of the absorption made measurements with the prisms impossible at wavelengths less than 0·64μ. Only one measurement of refractive index was made at 0·64μ and one at 0·65μ.

The results obtained by previous workers are also included in Figure 2. The present results agree to within less than 1% with the results of Wood and of Merwin and Larsen in the region of overlap, obtained by the same method.

---

* This amplifier was designed and developed by W. P. Melling, late of this Laboratory.

† These results were obtained from the *t* derived statistic whose probability distribution is known as the 'Student's' distribution. For a detailed account see *First Course in Mathematical Statistics*, by C. E. Weatherburn, chapter 10.

The results obtained by the polarimetric method used by Meier do not agree with those of Wood in the corresponding region.

(ii) *Absorption Coefficient of Amorphous Selenium*

The absorption coefficient of amorphous selenium was determined in the visible spectral region using selenium films evaporated in vacuum on to glass slides. The film thicknesses were determined by weighing the glass plates before and after deposition of the film. These thickness determinations assumed that the films were uniform and that their density was equal to the bulk density. Gmelin Institut (1942) quotes values of density between 4·26 and 4·302 gm/cm^3. The value used in this work was 4·3 gm/cm^3.

For measurements in the visible region a Hilger Barfit constant-deviation spectrometer was used. The detecting system was a Cintel VA 39 photometric cell and balanced cathode follower circuit. In the far red region a Cintel VS 39 photocell was used. The films under investigation were placed between the light source and the entrance slit of the spectrometer.

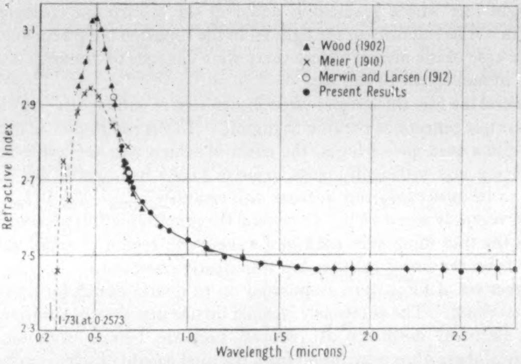


Figure 2. Comparison of results for the refractive index of amorphous selenium.

The photocell and cathode follower were used to measure the transmitted intensity for each film in turn at a given wavelength. The light sources used were a 250-watt high-pressure mercury arc lamp (B.T.H. type ME/D) and a 6-volt 18-amp. tungsten filament lamp supplied from a constant-voltage transformer. The film thicknesses varied from 40μ to 8μ in the far red region and from 0·5 to 0·05μ in the violet and near ultra-violet region. A total of forty films were used in this work. For the thinner films the thicknesses were determined from the total weight of selenium evaporated, the ratio of mass of selenium evaporated to mass of selenium on the plate having previously been determined for thicker films. The glass plates upon which the films were evaporated were first cleaned with Teepol, distilled water and ether. The plates were finally polished with a clean Selvyt cloth. When not in use the films were kept in a desiccator.

In the transmission measurements in the visible any scattered red light in the spectrometer would seriously affect the results owing to the low absorption coefficient of selenium for this spectral region. Appropriate filters (Chance

Bros.) were placed between the selenium film and the entrance slit of the spectrometer to minimize the effect of the scattered light. The spectral sensitivity of the VA 39 photometric cell was particularly suitable for this work as its sensitivity decreased rapidly for wavelengths greater than 0·60μ, in which region any scattered red light from the spectrometer would be readily transmitted by the selenium. Preston's (1936) method of correcting for scattered light in the spectrometer—whereby one covers half the entrance slit and half the exit slit—was tried at a wavelength of 0·50μ and the correction was found to be negligible. The slit range of the spectrometer varied from 60 A. at 0·75μ to 10 A. at 0·40μ.

From the measurements the absorption coefficient could be calculated in the wavelength region 0·75 to 0·365μ. Provided the incident intensity I_0 remains constant and the film thickness t is not less than 100 A. or so, a graph of the logarithm of transmitted intensity I plotted against t at a constant wavelength is a straight line whose gradient is equal to the absorption coefficient at that wavelength. The intensity of the light from the tungsten lamp remained constant but in the case of the mercury lamp there were changes of intensity of from 10 to 15% due to movement of the arc.

In general the film thicknesses were chosen to give sufficiently large absorption so that multiple reflections became negligible. In the red region of the spectrum the thick films used gave fringes, the effect of which was negligible because the fringe spacing was sufficiently close, relative to the bandwidth, for the average intensity to be observed, and because the visibility $(I_{\max}-I_{\min})/(I_{\max}+I_{\min})$ of the fringes was only about 15%. In general the graphs (log I, t) were straight lines, but when the thin films were used in a wavelength region of small absorption a deviation from the linear relationship was clearly observed.

A further set of films were evaporated on to quartz plates for measurements in the ultra-violet. The previously cleaned quartz plates were subjected to three minutes' electrical discharge at reduced pressure before evaporation. The thicknesses of these films were found from the total weight of selenium evaporated. The thickness varied from 0·03 to 0·5μ. Transmission measurements were made on these films for wavelengths between 0·4046 and 0·22μ. For this work a Hilger Barfit ultra-violet spectrometer was used together with a Cintel QVA 39 photometric cell and balanced cathode follower circuit. The light source used was a 125-w. mercury arc lamp (G.E.C. type MB/V).

Owing to the magnitude of the absorption and the lack of an accurate knowledge of the thicknesses of the films these experiments in the ultra-violet are rather inconclusive. There were indications of the absorption coefficient having reached a maximum at a wavelength of around 0·27μ but this is not known for certain.

The thicknesses of some of the films used have since been determined* using Tolansky's fringes of equal chromatic order. These preliminary measurements were made on four films in the thickness range 1·8 to 7·1μ and on three films in the range 0·031 to 0·111μ. The differences between the thicknesses obtained by Tolansky's method and those by weighing were found to be less than 10% for the thicker films and less than 20% for the thinner films.

The overall accuracy of the absorption measurements is difficult to assess but, taking into account the uncertainty of film thickness, fluctuations in the intensity of the mercury lamp and slight errors due to interference, the results over the five orders of magnitude measured are probably correct to within about 30%.

---

* By P. D. Fochs of this Laboratory.

Attempts were made to measure the absorption coefficient of the amorphous material in the near infra-red. As the absorption in this region is very small, cast blocks of selenium were used. It was not found possible, however, to make blocks sufficiently free of crystallinity for these measurements. Work on this problem is still continuing.

The results obtained for the absorption coefficient are plotted, together with those of previous workers, on a logarithmic scale against frequency in Figure 3.

(iii) *Absorption Coefficient of Crystalline Selenium*

In an endeavour to locate the absorption edge in the crystalline material measurements were made of the transmission of three single crystals of selenium*. The thicknesses of these crystals, found by weighing and on the assumption of a uniform thickness and of a density of 4·8 gm/cm^3, were 13·5μ, 46μ and 88μ. The 46μ crystal was the largest and most uniform. Its area was approximately 15 mm. × 1·5 mm. Accurate measurements of absorption with the crystalline material were severely limited by (i) the lack of uniformity of the crystals, (ii) their small size, (iii) their relatively large thickness and (iv) unknown loss of transmitted intensity due to reflection and surface scattering.

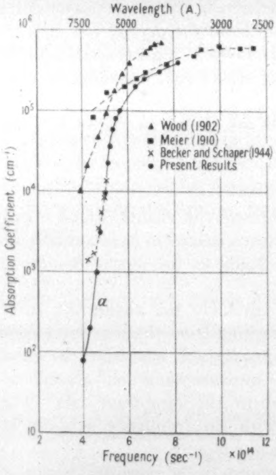


Figure 3. Comparison of results for absorption coefficient of amorphous selenium.

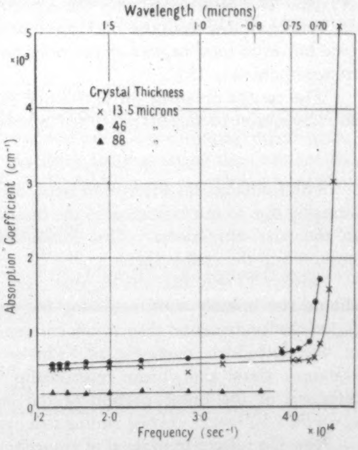


Figure 4. Absorption coefficient of crystalline selenium.

Each crystal was mounted in the aperture of a blackened brass block. The intensity of the transmitted light was measured with and without the crystal in the aperture. Care was taken to ensure that the crystal occupied the whole of the aperture in the mount.

---

* These crystals had been grown from the vapour phase by Dr. K. W. Plessner, late of this Laboratory.

The Hilger Barfit spectrometer was used to monochromatize the light; the detector was a lead sulphide cell whose signal was fed into the homodyne amplifier. A Cintel photometric cell and balanced cathode follower were also used for some of the intensity measurements.

In calculating the absorption coefficient allowance has to be made for losses in intensity due to reflection at both faces of the selenium crystal. In calculating the reflection coefficient for the crystalline material, its refractive index was assumed to be 1·2 times greater than that for amorphous selenium. This ratio of the refractive indices of the two forms of selenium was found by Gripenberg (1913). The values of the refractive index of amorphous selenium used for the calculations were those obtained in the present work (§ 3 (i)).

In the infra-red where the transmission is high, the correction, to be made due to reflection, is comparable with $\log I_0/I$, whereas in the visible where $\log I_0/I$ is large the correction is considerably less.

Twenty-two series of measurements were made of the variation of transmission coefficient with wavelength. With the $13{\cdot}5\mu$ crystal difficulty was found in measuring the ratio of incident to transmitted intensity in the visible ($I_0/I \sim 2{,}000$). To overcome this difficulty when measuring the incident intensity, a $0{\cdot}45\mu$ film of amorphous selenium on glass was interposed in the system. From the previous measurements of the absorption coefficient of amorphous selenium in this region one could determine the transmission of the crystal. It was assumed that the reflection loss from the crystal was equal to the reflection loss from the selenium film and glass base. This is approximately true but even total neglect of the reflection term would only introduce an error of approximately 15%.

The results obtained for crystalline selenium are shown in Figure 4, where the absorption coefficient is plotted against frequency.

§ 4. DISCUSSION OF RESULTS

The fundamental absorption band of selenium in the visible, which is presumably due to the excitation of the bound electrons, seems to have a maximum in the near ultra-violet. This excitation also leads to the strong dispersion observed in the visible region.

Seiwert (1949) has shown that very approximately the absorption curve should rise linearly with increasing frequency starting from the long-wave edge (ν_0) of the fundamental absorption, corresponding to $\boldsymbol{h}\nu_0 = E$, where E is the width of the forbidden energy gap. The present measurements on amorphous selenium show this linear relationship except in the long-wave tail. The intercept of the linear portion of the curve with the frequency axis lies at $\nu_0 = 4{\cdot}96 \times 10^{14}\,\mathrm{sec}^{-1}$, corresponding to a band spacing of 2·05 ev.

Now the refractive index n of amorphous selenium in the infra-red, assuming no further absorption band in the longer wavelength region, is related to the low-frequency dielectric constant ϵ_0 by the relation $n^2 = \epsilon_0$.

From the measurements of refractive index $n^2 = 6{\cdot}05$, whereas measurements of the D.C. dielectric constant indicate a value of 6·31 (Tammann and Boehme 1931). This leads one to expect a further absorption band in the infra-red.

In the long-wave region beyond the absorption edge (part *a* of Figure 3) the absorption coefficient drops very rapidly and approximately obeys a relation of the form $\mu = Ae^{B\nu}$, where $A = 8{\cdot}3 \times 10^{-9}\,\mathrm{cm}^{-1}$ and $B = 5{\cdot}5 \times 10^{-14}\,\mathrm{sec}$.

It was not possible to take measurements sufficiently far into the shorter wavelength range in the case of the crystalline samples to make an accurate estimate of the band spacing. The band spacing would appear, however, to be slightly less than that for the amorphous material. The chief contrast between the amorphous and crystalline material occurs in the long-wave tail. In the case of the crystalline material the absorption coefficient in the long-wave is fairly constant at several hundred cm^{-1}, whereas for the amorphous material it is 80 cm^{-1} at $0{\cdot}75\mu$ and, from preliminary measurements on the bulk material, is less than 1 cm^{-1} at a wavelength of 2μ. The higher absorption in the long-wave tail for the crystalline material is probably due to internal scattering at imperfections in the crystal lattice.

Interest in the shape of the absorption curve for amorphous selenium has lately increased since Weimar (1950) has found indications that amorphous selenium is a photoconductor possessing properties markedly different from those of either the metallic form or the red monoclinic crystals. Weimar found that the spectral response is generally peaked in the blue–green portion of the spectrum and is very small in the red region of the spectrum.

According to Seitz (1938) photoconductivity is not necessarily connected with the fundamental absorption but rather with some other absorption mechanism not yet fully understood. Now Weimar claims quantum efficiency approaching unity in the peaked region of the spectral response. Since this peaked region lies well within the fundamental absorption band it would appear that, in the case of amorphous selenium at least, the photoconductivity is directly connected with the fundamental absorption.

ACKNOWLEDGMENTS

The author is indebted to Mr. W. M. Jones of this Laboratory for developing much of the technique and for carrying out some of the preliminary experiments. He wishes to express his thanks to Dr. E. Billig for suggesting the problem and for many helpful discussions, and to Dr. T. E. Allibone for permission to publish this paper.

REFERENCES

BECKER, A., and SCHAPER, I., 1944, *Z. Phys.*, **122**, 49.

GMELIN INSTITUT, 1942, *Handbuch der Anorganischen Chemie*, 8 Aufl., System Nummer 10, Selen, Teil A, (Verlag : Chemie GMBH), p. 179.

GRIPENBERG, W. S., 1913, *Phys. Z.*, **14**, 123.

MEIER, W., 1910, *Ann. Phys., Lpz.*, **31**, 1017.

MERWIN, H. E., and LARSEN, E. S., 1912, *Amer. J. of Sci.*, **34**, 42.

MILLER, R. F., 1925, *J. Opt. Soc. Amer.*, **10**, 621.

PRESTON, J. S., 1936, *J. Sci. Instrum.*, **13**, 368.

SEITZ, F., 1938, *Faraday Society Discussion on Luminescence*, p. 98.

SEIWERT, R., 1949, *Ann. Phys., Lpz.*, **6**, 241.

SKINNER, C. H., 1917, *Phys. Rev.*, **9**, 148.

TAMMANN, G., and BOEHME, W., 1931, *Z. Anorg. Chemie*, **197**, 1.

WEIMAR, P. K., 1950, *Phys. Rev.*, **79**, 1, 171.

WELD, R. D., 1922, *J. Opt. Soc. Amer.*, **6**, 67.

WOOD, R. W., 1902, *Phil. Mag.*, **3**, 607.

On a New Test Method for Spherical Aberration of Electron Lenses

By O. KLEMPERER
Imperial College, London

MS. received 1st *December* 1950

ABSTRACT. The focus of rays from a lens with spherical aberration appears as a spot surrounded by a discrete halo ring if a diaphragm with fine circular aperture is placed across the beam in front of the focus. The diameter of the halo allows an estimate of the magnitude of the aberration involved. The geometry of rays forming the halo is explained here by schematic drawings. The practical application of the halo test is illustrated by some examples. In particular, the negative spherical aberration produced by an electronic space charge in a saddle field lens is demonstrated.

---

§ 1. INTRODUCTION

PRACTICAL methods for testing the spherical aberration of glass-lenses can be divided into three categories: (i) the 'Hartmann test', in which the beam is split up into pencils of given distances from the axis so that the focus for each aperture can be traced separately; (ii) the 'knife-edge test', in which a straight edge is placed across the beam and the aberration is derived from the shape of the shadow; (iii) the 'star test', in which the expanded out-of-focus image of a point source is examined on both sides of the best focus.

Methods of the first two categories have been adapted to electron optical requirements: adaptations of (i) are, for instance, the 'pepperpot-methods', by Epstein (1936) and by Klemperer and Wright (1939), and of (ii), the shadow methods, by Spangenberg and Field (1942) and by Dosse (1941), or the knife-edge method, by Liebmann (1949). On the other hand, the star test has not been adapted to the testing of electron lenses because of the difficulty of producing electron sources of sufficiently homocentric and isotropic emission.

In the following is described a new method for testing electron lenses for spherical aberration, to which the author has been led by the study of the ring focus in electron lens spectrometers (cf. DuMond 1949, Persico 1949, Slätis and Siegbahn 1949, Bothe 1950, Keller *et al.* 1950, Verster 1950).

In the new method, an apertured diaphragm intercepts the electron beam near the focus. If the focusing system is free from aberration the focus is observed as a fine spot on a fluorescent target behind the diaphragm. If, however, any axial aberrations are present, the focus appears to be surrounded by a discrete ring or 'halo'. This 'halo test' has proved to be useful for a quick recognition of the state of correction of an electron lens system.

§ 2. FORMATION OF A HALO IN PRESENCE OF ABERRATION

The formation of a halo ring about the focus may be explained with the help of Figure 1, which shows a meridional cross section through a circular beam. 0...0 represents the axis of the beam; 1, 2, 3, 4, 5 and 6 are rays which initially are equidistant and parallel to the axis. P is the principal plane of the system. The rays are shown to be refracted with positive spherical aberration, the marginal rays 6 being focused first, the paraxial rays 1 being focused last. A diaphragm D with a narrow, circular aperture is fixed at right angles to the axis in the plane containing the focus of the zonal rays 4. In some distance from the diaphragm D there may be arranged a fluorescent target T. Two discrete groups of rays can be seen to pass the aperture: (i) a paraxial group forming a well-defined spot on the target; (ii) a zonal group surrounding the rays 4, producing the picture of an annular ring on the target. The two regions illuminated by electron rays are marked in the figure by double lines parallel to the target. The rays 2 and 3 are intercepted by the diaphragm so that a dark zone is left between annular ring and spot. Also rays 5 and 6 are intercepted so that the area outside the bright ring is dark again. If the target T is gradually brought closer to the diaphragm D the sizes of ring and spot on the target decrease and become brighter. The diameter of the spot reaches a minimum at a disc of least confusion, but at a further decrease of the diaphragm-to-target distance it grows again, while the ring diameter further contracts until both pictures merge. A picture of the spot in minimum position surrounded by a bright ring (=halo) is shown in Figure 2(*a*).

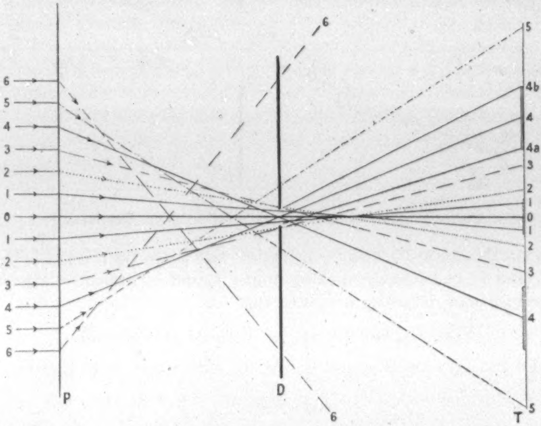
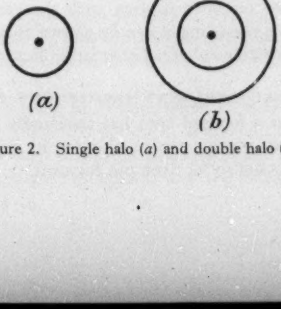


Figure 1. Halo formation in presence of aberration.


(*a*)

(*b*)

Figure 2. Single halo (*a*) and double halo (*b*).

Figure 3 illustrates the connection between the radii of the halo and the amount of the longitudinal spherical aberration involved. It shows on an enlarged scale the rays 4a and 4b which surround the rays 4 of Figure 1 and which are the critical rays just still passing through the aperture in the diaphragm D. The ray 4 may cut the axis at F_4, making an angle θ with it, the rays (4a) and (4b) cut the axis at F_a and F_b respectively, making angles γ_a and γ_b respectively with rays 4. If the aperture is of radius r_D, it follows that

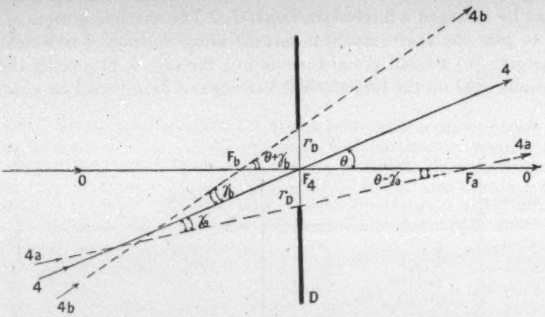


Figure 3. Radii of halo and amount of spherical aberration.

$$F_4F_a = r_D/\tan(\theta-\gamma_a); \qquad F_4F_b = r_D/\tan(\theta+\gamma_b).$$

For small θ and very small γ_a and γ_b, i.e. for sufficiently small apertures r_D, one obtains

$$F_aF_b \simeq 2r_D/\theta. \qquad \ldots\ldots(1)$$

The angle θ can be obtained as the ratio of the mean radius r_H of the halo and its distance from the diaphragm; moreover, the radius r_D of the aperture is known, so that, according to equation (1), the longitudinal aberration difference F_aF_b is accessible to measurement.

If the aberration follows approximately the usual square law, namely

$$F_nF_0 \simeq a\theta^2, \qquad \ldots\ldots(2)$$

the aberration coefficient a may be found as follows: $F_bF_0 = F_bF_a + F_aF_0$; hence

$$F_bF_a \simeq a(\theta-\gamma_b)^2 - a(\theta+\gamma_a)^2 \simeq -2a\theta(\gamma_a+\gamma_b)$$

or

$$a \simeq F_aF_b/4\gamma\theta \qquad \ldots\ldots(3)$$

where $\gamma = (\gamma_a+\gamma_b)/2$ is the semi-aperture of the annular pencil. γ can be obtained by measuring the width of the annular ring on the target, the diameter of the aperture and the distance between target and aperture.

The discussion has so far been dealing with the case of positive spherical aberration. By analogous reasoning it can be shown that halo formation occurs in a similar way when negative spherical aberration is present.

§ 3. PRACTICAL APPLICATION OF THE HALO TEST

Halo formation about a focused spot has repeatedly been described in the electron optical literature, but an explanation for it, based on the geometrical aberration of the rays, has not so far been put forward.

Myers (1939), for instance, has reproduced a picture of a sharp halo about a point image which he ascribes to secondary emission from the edge of the anode aperture. In a space-charge modulated hexode gun, described by Shoenberg, Condliffe and Tedham (1933) and widely applied in early television practice, halo formation about the focus was found to be quite disturbing as soon as the beam current was partially biased off. It may be suggested that the halo formation in this hexode is due to spherical aberration of the lens formed in the modulator electrode.

Some quantitative examples for the application of a halo test to some simple electron lenses may be added here. For the test the lenses should focus rays from an aberration-free electron source through a fine aperture. As such a source may be used, for instance, a hairpin emission system which, according to recent measurements by Klemperer and Klinger (1951), is known to be reasonably free from aberration. Now either the focal length of the electron lens or the distance of aperture and target from the lens may be adjusted until the halo ring reaches a maximum radius r_H, while its breadth Δr_H simultaneously reaches a minimum. Some results are listed in the Table.

Lens	No. 1		No. 2	No. 3			
V'/V	5·0	5·0	4·6	3·2	3·2	3·2	3·2
MJ	95	95	95	75	75	75	75
MT	335	335	335	355	355	355	355
DT	175	175	175	175	175	175	175
r_D (mm.)	0·025	0·28	0·28	0·28	0·28	0·025	0·025
r_H (mm.)	4·5	5·5	5·0	3	7	2·5	6
Δr_H (mm.)	0·3	0·8	0·1	0·4	0·6	0·15	0·10
θ	$2{\cdot}5\times10^{-2}$	$3{\cdot}2\times10^{-2}$	$2{\cdot}9\times10^{-2}$	$1{\cdot}7\times10^{-2}$	$4{\cdot}0\times10^{-2}$	$1{\cdot}4\times10^{-2}$	$3{\cdot}4\times10^{-2}$
γ	$0{\cdot}8\times10^{-3}$	$2{\cdot}1\times10^{-3}$	$0{\cdot}3\times10^{-3}$	$0{\cdot}8\times10^{-3}$	$1{\cdot}5\times10^{-3}$	$0{\cdot}4\times10^{-3}$	$0{\cdot}2\times10^{-3}$
F_aF_b (mm.)	2	17	19	33	14	3·6	1·5
a/R	2×10^3	5×10^3	4×10^4	5×10^4	4×10^3	$1{\cdot}3\times10^4$	4×10^3

V'/V = voltage ratio; MJ = distance (mm.) of point source J from midplane M of lens; MT = midplane-to-target distance (mm.); DT = diaphragm-to-target distance (mm.).

The lens numbers are those of the types of lenses under investigation; in particular, No. 1 is a symmetrical two-tube lens of 12·7 mm. tube radius, accelerating the electron beam; No. 2 is a decelerating lens with the same electrodes; No. 3 is a symmetrical saddle field lens consisting of three tubes of 12·7 mm. radius, the intermediate tube being three tube radii long and at less positive potential than the external tubes. The other symbols are as defined in the text and by Figure 3. F_aF_b is the distance between foci of rays of different aperture obtained according to equation (1). The angles θ and γ are either found approximately as the ratios $\theta = r_H/\mathrm{DT}$ and $\gamma = \Delta r_H/2\mathrm{DT}$ or, more accurately, they are traced with a sliding, fluorescent target.

F_aF_b represents the longitudinal spherical aberration difference for two beams of angular aperture $\theta + \gamma$ and $\theta - \gamma$, hence the magnitude of F_aF_b depends upon θ, γ and on the optical quality of the lens employed. Thus it is clear that larger F_aF_b should result from larger semi-apertures r_D; this is borne out in the Table, as F_aF_b increases from 2 to 17 mm. when r_D is increased from 0·025 to 0·28 mm. On the other hand, the spherical aberration coefficient a calculated according to equation (3) and measured in units of the tube radius R (bottom line of Table) should be largely independent of the particular values of θ, γ and

F_aF_b. This is confirmed only approximately by the measured results. The observed variation of a/R by a factor 2 may be expected from inaccuracies of measurement as well as from the fact that the square-law equation (2) applies only as a rough approximation to the lenses listed in the Table.

However, the values obtained for a/R are a measure of the quality of a lens. According to the results obtained by other methods (cf. Klemperer and Wright 1939), a/R should be much larger for a decelerating two-tube lens (No. 2) than for a corresponding accelerating lens (No. 1). This is again in agreement with the results given in the Table, which shows the aberration coefficient a/R of lens No. 2 to be about ten times larger than that of lens No. 1.

Of particular interest are the results of the halo test applied to a three-tube saddle field lens with relatively long intermediate tube electrode, the potential of which is negative with respect to the outer electrodes. Klemperer (1939) has shown that the space charge of the electron beam passing through such a lens sets up a negative spherical aberration for the rays of small aperture. On the other hand, rays of larger aperture are not expected to traverse the space-charge cloud, since apparently this is formed in the paraxial region only. Hence marginal rays would show the usual positive spherical aberration. Now the halo test has been found to respond to both aberrations. A double halo consisting of two discrete rings round the spot as shown in Figure 2 (*b*) has been observed on the fluorescent target. The ring of smaller radius r_H belongs to the rays of smaller aperture, and apparently it is caused by the space charge. As seen from the values of a/R in the Table, the aberration coefficient due to space charge is very much larger than that due to geometrical optical lens error. Moreover, the space-charge aberration apparently deviates very much from the square law of equation (2), as follows from the relatively large divergence of results obtained with diaphragm apertures of different radius.

These examples should be sufficient to illustrate the use of the halo method for quick testing of electron lenses. Moreover, they may demonstrate its value in the detection of aberrations of rays of very small apertures. For obtaining exact quantitative information on the whole aberration curve the halo test can so far not compete with the established standard methods.

REFERENCES

Bothe, W., 1950 a, *S.B. Heidelberg Akad. Wiss.* (*Math.-Naturwiss. Kl.*), 191; 1950 b, *Naturwissenschaften*, **37,** 41.

Dosse, J., 1941, *Z. Phys.*, **117,** 722.

DuMond, J. W. M., 1949, *Rev. Sci. Instrum.*, **20,** 160.

Epstein, D. W., 1936, *Proc. Inst. Radio Engrs.*, *N.Y.*, **24,** 1095.

Keller, J. M., Koenigsberg, E., and Paskin, A., 1950, *Rev. Sci. Instrum.*, **21,** 713.

Klemperer, O., 1939, British Patent Specification No. 534.215.

Klemperer, O., and Klinger, Y., 1951, *Proc. Phys. Soc.* B, **64,** 231.

Klemperer, O., and Wright, W. D., 1939, *Proc. Phys. Soc.*, **51,** 296.

Liebmann, G., 1949, *Proc. Phys. Soc.* B, **62,** 213.

Myers, L. M., 1939, *Electron Optics* (London: Chapman and Hall).

Persico, E., 1949, *Rev. Sci. Instrum.*, **20,** 545.

Shoenberg, I., Condliffe, G. E., and Tedham, W. F., 1933, British Patent Specification No. 431,327.

Slätis, H., and Siegbahn, K., 1949, *Ark. Fys.*, **1,** 339.

Spangenberg, K., and Field, L. M., 1942, *Proc. Inst. Radio Engrs.*, *N.Y.*, **30,** 138.

Verster, N. F., 1950, *App. Sci. Res.* B, **1,** 363.

The Mechanism of Positive Ion Collection by a Spherical Probe in a Dense Gas

By R. L. F. BOYD
Department of Physics, University College, London

Communicated by H. S. W. Massey; *MS. received* 15*th February* 1951

ABSTRACT. The Langmuir probe technique is not suitable for measuring ion densities above 1 mm. Hg pressure because the probe dimensions approach those of the ionic and electronic mean free paths. In addition to the invalidity of the Langmuir theory for this case, such a probe also causes a great disturbance of the discharge.

In this paper a detailed examination is made of the possibility of using a probe collecting positive ions as a means of finding ion densities.

It has been found possible to calculate the potential distribution, outside a space-charge sheath around a spherical probe, if the radius of the sheath is known. From this the current of positive ions to the probe may be found. Curves are given to facilitate this calculation.

It is found that, depending on the ion concentration and pressure, there are two rather different sets of conditions around the probe. If the ion density is high (greater than about 10^8/mean free path (in cm.)) a thin sheath will form on the probe and a complete solution of the problem is possible. Under such conditions the energies of the diffusing ions in the neighbourhood of the probe greatly exceed the thermal energies of the gas particles. At the lower ion densities normally encountered, a thick 'diffusion' sheath occurs and the problem can now only be solved if the sheath thickness is known. The ion energies may or may not greatly exceed the thermal energies, depending on conditions.

These results contradict one of the basic assumptions made by Davydov and Zmanovskaja in their approach to the same problem. They assume a thin sheath with the ions in thermal equilibrium with the surrounding gas. As a result their conclusion is in conflict with the conclusion of this paper, that in most circumstances it is not possible to use a negative probe to measure ion densities unless an additional means of determining the sheath thickness is available.

§ 1. INTRODUCTION

In the study of the gas discharge, pressures greater than 1 mm. Hg are commonly encountered, and the problem of determining the ionization density frequently recurs. The purpose of this paper is to examine the suitability of a probe operating at a potential somewhat negative to the space as a means of determining this quantity. The use of Langmuir's probe technique is limited to pressures at which the mean free path of the electrons is large compared with the probe diameter (be it disc, cylinder or sphere); if the probe be cylindrical, a diameter of 10^{-3} cm. is about the lowest useful limit. For spherical and plane probes the limit is much higher. This means that at pressures above 1 mm. Hg the Langmuir technique is no longer available as a means of measuring ion density. Moreover, even below 1mm. Hg the satisfactory operation of very fine wire probes depends upon the possibility of allowing for the positive ion component of current to the probe by extrapolating the 'saturation' part of the positive ion characteristic. Such an extrapolation, which is never very accurate, for reasons set forth in a recent paper (Boyd 1950, to be referred to as I), is meaningless when the diameter of the space-charge sheath surrounding the probe under the saturation conditions is much larger than the probe diameter. To crystallize our ideas we note that for a probe potential of 20 volts and an ion density of $10^{12}\,\mathrm{cm}^{-3}$ under normal discharge conditions the sheath thickness is about 10^{-2} cm. At a density of $10^{10}\,\mathrm{cm}^{-3}$ it is about 10^{-1} cm.

The idea of using a probe large compared with the mean free path of the ions and attempting to solve the equations defining the flow of current to the probe is attractive. Such an attempt has been made by Davydov and Zmanovskaja (1936), and some criticism of their treatment will be given here. These authors give results relating to a wide range of probe potentials, positive and negative. For practical usefulness, however, only a negative probe need be considered since in a dense gas the increased ionization rate in the probe neighbourhood, when electrons are being accelerated towards the probe, vitiates the theory. Moreover the drain of current from the discharge under such conditions is liable to be a severe disturbance, being two or three orders of magnitude greater than under negative probe conditions.

For this reason the following treatment will be confined to the case of an electronegative probe, in an effort to find under what conditions the measured positive current to such a probe can give information concerning the ion density. It will be well to anticipate the results here by remarking that the conditions will be found to be very stringent, much more so than is apparent in the treatment of Davydov and Zmanovskaja.

By taking the case of a spherical probe results can be obtained which are independent of the discharge-tube radius. This is not the case for a cylindrical probe. Moreover, the difficulty encountered in I in dealing with the blending of the diffusion region with the undisturbed plasma for a small plane probe is absent in the case of a spherical probe. The effect of a small lead and support for the sphere is neglected.

An interesting feature of the results will be seen to be the dependence of the positive-ion current upon the electron temperature rather than on the ion velocities. The electron temperature can of course be measured at these pressures from the slope of the (log current, voltage) curve for a probe at fairly high retardation voltages (I).

§ 2. DEFINITIONS

(It is convenient to introduce certain dimensionless parameters to represent position and potential.)

a_0 = probe radius; a_1 = outer radius of sheath; a_2 = outer radius of abnormal diffusion region; a_3 = outer radius of normal diffusion region.

r = radial distance from an origin at the centre of the probe; $s_0 = r/a_0$, the length parameter; $s_1 = r/a_1$ etc.

λ = ionic mean free path; $l = \lambda/a_1$.

p = gas pressure; $\beta = \lambda p$, a constant for a given gas.

V_{p} = probe potential; V = potential, V_0 at position s_0, V_1 at s_1 etc.; $V_{\mathrm{e}} = \boldsymbol{k} T_{\mathrm{e}}/\boldsymbol{e}$ = electron temperature (in ev.); it is $\frac{2}{3}$ the mean energy of the electrons; $\boldsymbol{k}$ = Boltzmann's constant; $\boldsymbol{e}$ = magnitude of electronic charge; T_{e} = absolute electron temperature; $V_{\mathrm{g}} = \boldsymbol{k} T_{\mathrm{g}}/\boldsymbol{e}$ = gas temperature, where T_{g} = absolute gas temperature.

$E = V/V_{\mathrm{e}}$, $E_0 = V_0/V_{\mathrm{e}}$ etc.; $E_{\mathrm{g}} = V_{\mathrm{g}}/V_{\mathrm{e}} = T_{\mathrm{g}}/T_{\mathrm{e}}$; $E_{\mathrm{p}} = V_{\mathrm{p}}/V_{\mathrm{e}}$.

η = ionic mobility at unit pressure for low field strength defined by $v_+ = (\eta/p)(dV/dr)$, where v_+ = positive ion radial drift velocity.

μ = ionic 'mobility constant' for high field strengths defined by $v_+ = (\mu/p^{1/2})(dV/dr)^{1/2}$.

n_+^s, n_+^∞ = concentration of positive ions at radial position s, and at infinity; n_{e}^s, n_{e}^∞ = concentration of electrons at radial position s, and at infinity; $\delta = (n_+^s - n_{\mathrm{e}}^s)/n_+^s$.

i_+ = positive ion current to the probe; j_+ = positive ion current density at the sheath edge ($s_1 = 1$).

The current of positive ions to a negative probe in a discharge plasma at pressures for which $l \ll 1$ is determined by the potential distribution around the probe. In considering this distribution we distinguish four regions : (1) Between a_0 and a_1 is a region of strong field in which n_+ greatly outnumbers n_e (i.e. $\delta \sim 1$). This region is known as the 'sheath'. (ii) Between a_1 and a_2 there may be an intermediate region of strong field in which $n_+ \simeq n_e$ (i.e. $\delta \ll 1$). This region is referred to as the 'abnormal extra-sheath region'. (iii) Between a_2 and a_3 is a region of weak field in which also $n_+ \simeq n_e$. This is referred to as the 'normal extra-sheath region'. (iv) Beyond a_3 is a region in which the local disturbance due to the probe is 'lost' in the general discharge condition. This occurs either when the field of region (iii) falls to a value comparable with the normal discharge fields or when the extent of region (iii) becomes so great that it is no longer possible to neglect the ion production within it compared with the ion flux through it. This is called the 'undisturbed region'.

In the normal sheath region the flow of positive ions is governed by a mobility equation of the form $v_+ = (\eta/p)(dV/dr)$, but in the abnormal sheath region the flow is better represented by an equation of the form $v_+ = (\mu/p^{1/2})(dV/dr)^{1/2}$ (Sena 1946).

The importance of using Sena's mobility equation lies in the fact that the ordinary mobility equation is based on the assumption that the ions are in approximate thermal equilibrium with the gas. It has been shown in I however that conditions at the sheath edge—in particular whether a 'free-fall' sheath will form—are strongly dependent upon the energy distribution of the ions at the edge.

Figure 1 gives a diagrammatic representation of the potential distribution in the various regions, together with the conditions prevailing at their boundaries. (These are derived later.)

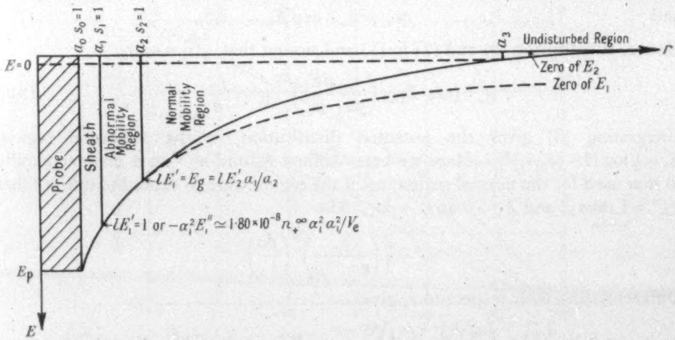


Figure 1. Distribution of potential around a spherical probe (diagrammatic).

§ 3. THE DISTRIBUTION OF POTENTIAL OUTSIDE THE SHEATH

(i) *The Normal Region*

The positive ion current current density at the sheath edge is $j_+ = i_+ / 4\pi a_1^2$ so that

$$n_+^{\,s} v_+^{\,s} s_2 = \frac{-j}{\mathbf{e}} \left(\frac{a_1}{a_2}\right)^2. \qquad \ldots\ldots(1)$$

If we assume the current of electrons to the probe is small compared with their random current (i.e. if $E_p \gtrsim 4$) the electron density is given by Boltzmann's relation $n_e^s = n_e^\infty \exp E_2$, where n_e^∞ is the limit to which the density of the electrons would approach at infinity, i.e. $n_e^\infty = \lim_{s_2 \to \infty} n_e^s$. The zero of potential for this region is defined by $E_2 \to 0$ as $s_2 \to \infty$.

Now in this region

$$v_+^s = \frac{-\eta}{p} \frac{dE_2}{ds_2} \frac{V_e}{a_2} \qquad \ldots\ldots(2)$$

and $n_+^s \simeq n_e^s$, so

$$n_+^s = n_+^\infty \exp E_2, \qquad \ldots\ldots(3)$$

and, therefore, substituting (2) and (3) in (1),

$$n_+^\infty (\exp E_2) \frac{\eta}{p} \frac{dE_2}{ds_2} \frac{V_e}{a_2} s_2^2 = \frac{j_+}{\mathbf{e}} \left(\frac{a_1}{a_2}\right)^2. \qquad \ldots\ldots(4)$$

Integrating (4),

$$E_2 = \log(1 - \alpha_2/s_2), \qquad \ldots\ldots(5)$$

where $\alpha_2 = j_+ p a_1^2 / \mathbf{e} n_+^\infty \eta V_e a_2$. This is the potential distribution in the normal region with n_+^∞ and E_2 defined as for an infinite region. We require also to know $E_2'(= dE_2/ds_2)$ and $E_2''(= d^2E_2/ds_2^2)$; differentiating (5) with respect to s_2 $\alpha_2 E_2' = (1 - \alpha_2/s_2)^{-1} (\alpha_2/s_2)^2$ and $\alpha_2^2 E_2'' = -\alpha_2 E_2' (\alpha_2 E_2' + 2\alpha_2/s_2)$.

(ii) *The Abnormal Region*

In this region

$$v_+^s = -\mu \left(\frac{V_e}{pa_1} \frac{dE_1}{ds_1}\right)^{1/2}, \qquad \ldots\ldots(6)$$

and

$$n_+^s = n_+^\infty \exp E_1. \qquad \ldots\ldots(7)$$

Hence, substituting (6) and (7) in (1) and noting that $s_2/s_1 = a_1/a_2$,

$$n_+^\infty (\exp E_1) \mu \left(\frac{V_e}{pa_1} \frac{dE_1}{ds_1}\right)^{1/2} s_1^2 = \frac{j_+}{\mathbf{e}}. \qquad \ldots\ldots(8)$$

Integrating (8) gives the potential distribution in the abnormal region: $E_1 = \frac{1}{2} \log\{1 - (\alpha_1/s_1)^3\}$. Here we have defined E_1 and n_+^∞ in a manner parallel to that used for the normal region, i.e. if the region were to extend to infinity then $n_+^\infty = \lim_{s_1 \to \infty} n_+^s$ and $E_1 \to 0$ as $s_1 \to \infty$. Also

$$\alpha_1^3 = \frac{2}{3} \left(\frac{j_+}{\mathbf{e} n_+^\infty \mu}\right)^2 \left(\frac{pa}{V_e}\right).$$

Differentiating with respect to s_1 gives

$$a_1 E_1' = \frac{3}{2} \left\{1 - \left(\frac{\alpha_1}{s_1}\right)^3\right\}^{-1} \left(\frac{\alpha_1}{s_1}\right)^4; \quad \alpha_1^2 E_1'' = -2\alpha_1 E_1' \left(\alpha_1 E_1' + 2\frac{\alpha_1}{s_1}\right).$$

The potential distributions for the normal and abnormal regions are plotted in Figure 2 and the values of $\alpha E'$ and $\alpha^2 E''$ in Figure 3.

(iii) *Boundary Conditions for the Regions*

The transition between the regions will now be examined.

At the junction of the sheath with the abnormal diffusion region there are two possible conditions. If the ion density is very high the field may reach such a value that the sheath criterion derived in equation (7) of I is satisfied. Then a sheath is formed in which there is a voltage drop several times greater than the electron temperature in one mean free path (a free-fall sheath). This condition has been largely dealt with in I and receives further comment later in this paper. If the ion density is insufficient for the field to attain the criterion value without violation of the quasi-neutrality postulate then a diffusion sheath forms. Such a sheath may be many mean free paths thick for a change of potential equal to the electron temperature.

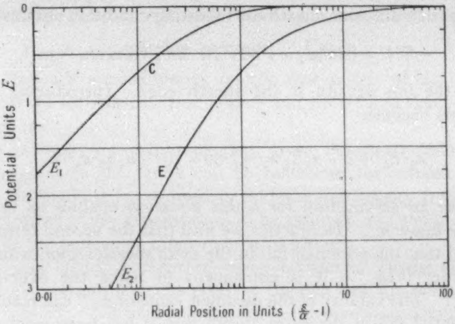


Figure 2. Potential distribution in the normal (E_1) and abnormal (E_2) mobility regions.

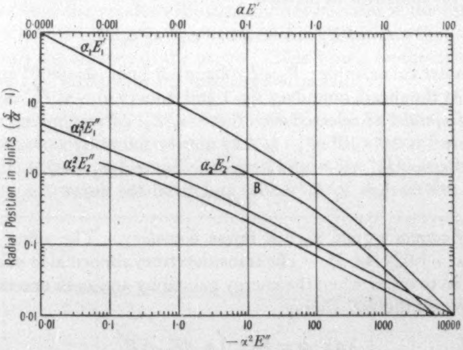


Figure 3. Distribution of the product of the electric field and its gradient by α and α^2 respectively for normal ($\alpha_1 E_1'$ and $\alpha_1^2 E_1''$) and abnormal ($\alpha_2 E_2'$ and $\alpha_2^2 E_2''$) regions.

It will be assumed for the present that the sheath thickness, and therefore the value of a_1, is known. The problem of determining a_1 is discussed in a later section.

(*a*) *The sheath, abnormal region boundary.* The value of the abscissa on the curves of E, E' and E'', which corresponds to the transition from abnormal region to sheath, is determined either by the fulfilment of the sheath criterion $\lambda\, dV/dr = V_e$ (i.e. $E_1' = 1/l$) or the failure of the quasi-neutrality postulate if, as is usually the case, this should occur at a lower value of field.

Obviously the failure of the quasi-neutrality postulate is a matter of degree. It seems reasonable to abandon the solution of the diffusion equations (4) and (8) and to assume that a diffusion sheath commences when δ has a value about 0·1. The choice of this value is necessarily somewhat arbitrary unless and until a satisfactory method can be found of defining the sheath edge and measuring its radius. At the higher ion densities, however (above 10^{12} cm^{-3}), the exact choice has very little effect upon the calculated current to the probe.

At the edge of a diffusion sheath the following condition obtains:

$$-\nabla^2 V = 4\pi \mathbf{e}\delta n_+ = 1{\cdot}80 \times 10^{-6}\delta n_+ \text{ volts cm}^{-1}\text{cm}^{-1}.$$

where n_+ is the ion density at the sheath edge. Introducing dimensionless parameters this becomes

$$\alpha_1^2 E_1'' + 2\frac{\alpha_1}{s_1}\alpha_1 E_1' = -1{\cdot}80 \times 10^{-6}\delta n_+ \alpha_1^2 a_1^2 / V_e. \qquad \ldots\ldots(9)$$

If we make an assumption for δ this equation enables us to find α_1 from Figure 2 if we know n_+. In practice we find that the second term in (9) can be neglected, and that the potential fall in the extra sheath region is such as to make n_+ of the order $0{\cdot}1 n_+^\infty$; it is convenient to make the overall assumption $\delta n_+ = 0{\cdot}01 n_+^\infty$. The validity of the assumed value of n_+^∞ can readily be checked once the potential fall in the extra sheath region has been worked out and an iterative correction made if necessary.

Equation (9) may now be written

$$\alpha_1^2 E_1'' \simeq 1{\cdot}80 \times 10^{-8} n_+^\infty \alpha_1^2 a_1^2 / V_e. \qquad \ldots\ldots(9\,a)$$

Consider a numerical example: $V_e = 2$ volts, $a = 0{\cdot}1$ cm., $\lambda = 0{\cdot}001$ cm., so $l = 0{\cdot}01$, $n_+^\infty = 10^{13}$. At the sheath boundary $s_1 = 1$ and from (9 *a*) $-\alpha_1^2 E_1'' \simeq 0{\cdot}9 \times 10^{-3}\alpha_1^2$.

A value of α_1 must be selected such that $-\alpha_1^2 E_1''$ corresponding to $1/\alpha_1$ on the curve of Figure 3 is $0{\cdot}9 \times 10^3 \alpha_1^2$. α_1 may now be found by a process of iteration. Try $\alpha_1 = 1$, then $\alpha_1^2 E_1'' = 900$ and from the curve $1/\alpha_1 = 1{\cdot}024$, i.e. $\alpha_1 = 0{\cdot}976$ now try $\alpha_1 = 0{\cdot}976$, then $\alpha_1^2 E_1'' = 860$, and from the curve $1/\alpha_1 = 1{\cdot}024$ and so $\alpha_1 = 0{\cdot}976$.

(*b*) *The abnormal region, normal region boundary.* The value of E_1 corresponding to $s/\alpha_1 = 1{\cdot}024$ is 1·35. The transition from abnormal to normal diffusion can be assumed to occur when the energy gained by an ion in one mean free path is equal to the gas thermal energy

$$\lambda\, dV/dr \simeq V_g, \quad \text{i.e. } lE_1' = E_g. \qquad \ldots\ldots(10)$$

Suppose $E_g = 0{\cdot}015$, then in the numerical example $E_1' = 1{\cdot}5$ at the transition. (This corresponds to $(dV/dr)/p = 30$ volts/cm/mm. Hg.)

Thus at s_1/α_1 corresponding to $\alpha_1 E_1' = 1{\cdot}5 \times 0{\cdot}976 = 1{\cdot}468$ the flow transfers to the normal region. This value of s_1/α_1 is 1·22 and corresponds to a value of $E_1 = 0{\cdot}4$. The potential fall in the abnormal region is thus $E = 1{\cdot}35 - 0{\cdot}40 = 0{\cdot}95$.

Now a_2/a_1 is equal to the value of s_1 at the boundary of these regions, i.e. $a_2 = a_1 \times 1{\cdot}22 \times 0{\cdot}976 = 1{\cdot}19\, a_1$. At the boundary $s_2 = 1$, so $s_2/\alpha_2 = 1/\alpha_2$. A value of α_2 must be selected such that at the junction of the two regions there is no discontinuity in the field strength: $E_1' V_e / a_1 = E_2' V_e / a_2$; in the example $E_2' = 1{\cdot}19 \times 1{\cdot}5 = 1{\cdot}78$.

By a process of iteration similar to that applied for the abnormal region it is found that $\alpha_2 = 0{\cdot}65$. The potential at $s_2/\alpha_2 = 1{\cdot}55$ is $E = 1{\cdot}05$ so that the total potential drop to the sheath edge is $E = 0{\cdot}95 + 1{\cdot}05 = 2{\cdot}0$. It is now possible to check the assumption $n_+ = 0{\cdot}1 n_+^\infty$ since $n_+ = n_+^\infty \exp(-E)$. It is found that $\exp(-E) = 0{\cdot}135$, which agrees sufficiently closely with the assumed value.

The effect on the value of α_1 and hence of α_2 of a much greater discrepancy in the assumed value of n_+ is readily seen from Figure 3 to be negligible owing to the closeness of α_1 to unity. Having found α_2, j_+ and i_+ can now be found for

$$i_+ = 4\pi a_1^2 j_+ = 4\pi \alpha_2 n_+^\infty \eta V_e a_2 / p. \qquad \ldots\ldots(11)$$

Note that i_+ depends approximately on the first power of a_2. Inserting the values $\eta = 2{,}100$ cm/sec/volt/cm. (for argon) and $\beta = 0{\cdot}01$ so that $p = 10$ mm. Hg, we find $i_+ = 6{\cdot}5$ ma.

From this example it can be seen that, given the sheath radius and the ion density, the current to the probe can be calculated for suitable values of the parameters. Normally it is required to find the ion density from the current. To do this a number of ion densities would be assumed and a graph of i_+ against n_+^∞ plotted for the particular experimental conditions.

(*c*) *The normal region, undisturbed region boundary.* In the above example, and indeed in general, no notice is taken of the fact that the normal region does not extend to infinity. Normally conditions in the plasma and its extent are such as to make this a valid approximation.

Care must be taken however if there is a very strong electric field in the discharge tube, and the results will certainly not be valid if the rate of ion production in a spherical volume having a radius only a few times that of the probe is comparable with the current to the probe. This would be the case if the probe were not small compared with the discharge-tube radius.

§ 4. PROCEDURE FOR FINDING i_+ FROM n_+

To obviate the need for iteration in finding α_1 and α_2, curves have been plotted in Figure 4 giving α_1 in terms of E_1'' at the sheath edge, α_2 in terms of E_2' at the normal–abnormal region boundary, α_2 in terms of E_2'' at the sheath edge.

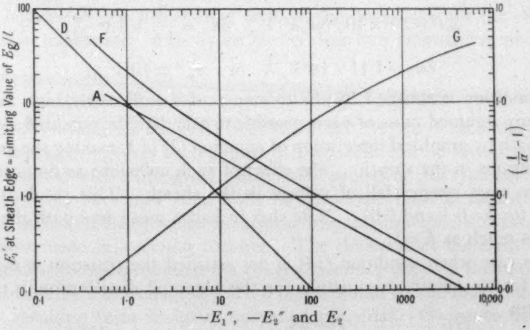


Figure 4. A is the value of α_1 in terms of E_1'' at the sheath edge, D the value of α_2 in terms of E_2' at the boundary and F the value of α_2 in terms of E_2'' at the sheath edge, all plotted in the form of the parameter $(\frac{1}{\alpha} - 1)$. G is the limiting value of E_g/l for which an abnormal diffusion region is present in terms of E_1'' at the sheath edge.

The last curve is employed when the abnormal region is absent, i.e. when a diffusion sheath commences before the field attains the value given by (10).

Curve G gives the value of E_1 at the sheath edge for various values of E_1 at the edge. If E_g/l exceeds this value the abnormal region is absent and it is only necessary to find α_2 by giving a_2 the value of the sheath radius.

The procedure for finding i_+ from n_+^∞ can be summarized thus: 1. Find α_1 from curve A and (9*a*). 2. Find E_1' at the transition from (10) (normally $E_g \simeq 0{\cdot}015$). 3. From curve B find s_1/α_1 corresponding to $\alpha_1 E_1'$ as found by 1 and 2. If $s_1/\alpha_1 < 1/\alpha_1$ (i.e. $s_1 < 1$) there is no abnormal region, so start again at step 1A below. 4. From curve C (Figure 2) find the difference in potentials corresponding to $1/\alpha_1$ from 1 and s_1/α_1 from 3. This is the potential fall in the abnormal extra sheath region. 5. Find a_2 in terms of a_1. Now $a_2/a_1 = s_1$ (at the boundary), so $a_2 = a_1(s_1/\alpha_1)\alpha_1$ as found from 3 and 1. 6. Find E_2' at the boundary such that $E_2' = E_1' a_2/a_1$ as found from 2 and 5. 7. Find α_2 from curve D and the value of E_2 from 6. 8. From curve E find the potential corresponding to α_2 from 7. This is the potential fall in the normal extra sheath region. 9. From the total potential fall E (sum of 4 and 8) find the ratio of the density at the sheath edge to that at ∞ from $n_+/n_+^\infty = \exp(-E)$. 10. Find the probe current i_+ from (11).

1A. Find α_2 from curve F and (9*a*) replacing E_1'' by E_2''. 2A. From curve E (Figure 2) find the potential corresponding to α_2 from 1A. This is the potential fall E in the extra sheath region. 3A. Proceed to 9 and 10 above.

§ 5. THE POTENTIAL DISTRIBUTION IN THE SHEATH

In I it has been shown that a free-fall sheath will form when $\lambda\, dV/dr = V_e$ (if the ion density is sufficient for a diffusion sheath not to have formed before), i.e. when

$$lE_1' = 1. \qquad \ldots\ldots(12)$$

Since l is small, it is clear from the curve of $\alpha_1 E_1'$ that s_1/α_1 is very close to unity under such circumstances. Substituting (12) in (9) and noting that $E_1'' \simeq -2/l^2 \simeq \nabla^2 E_1$, we have the condition that for a free-fall sheath, i.e. a sheath in which there are few collisions,

$$2/l^2 = 1{\cdot}8 \times 10^{-6} \delta n_+ a^2/V_e, \quad \delta n_+ = \frac{2}{1{\cdot}8} \times 10^6 \frac{V_e}{a^2 l}, \qquad \ldots\ldots(13)$$

$$\delta n_+ = 1{\cdot}11 \times 10^6 V_e/\lambda \quad \text{or} \quad n_+^\infty \simeq 10^8 V_e/\lambda. \qquad \ldots\ldots(14)$$

If this condition is satisfied by the existence of a sufficiently high ion density then for any assumed value of δ it is possible to calculate the potential distribution in the sheath by graphical integration of equation (2) of I, making the assumption of no collisions in the sheath. The effect of such collisions as do occur will be to give an even steeper fall of voltage in the sheath. This result is given in Figure 5 for $\delta = 0{\cdot}1$ and $0{\cdot}01$. Note that in half a mean free path the potential falls by as much as $E = 3$.

In the case when condition (14) is not satisfied the situation is much more difficult, for we obtain an equation for the potential distribution in the sheath of the form

$$\frac{d^2V}{dr^2} = -4\pi \boldsymbol{en} \left\{ \frac{C}{(dV/dr)^{1/2}} - (1-\delta) \exp\left(-\frac{V}{V_e}\right) \right\},$$

where C is some constant.

While an approximate solution can of course be obtained for any given set of parameters, no general solution is possible, and it would seem that the only practicable course is to make a subsidiary experimental determination of the sheath thickness. This is a severe limitation to the usefulness of what would otherwise seem to be a very promising method.

Three methods of measuring the sheath thickness suggest themselves: a cathetometer, the R.F. capacitance of the sheath and an auxiliary movable probe. None of these methods promises to be very satisfactory. A cathetometer might be used in certain cases, although there must be a fair degree of uncertainty as to the correspondence between the measured value for the dark sheath radius and the value of a_1 as defined by the assumed value of δ at the sheath edge. The R.F. method might be developed for fairly small sheaths, but for sheaths many times greater than the probe radius insufficient accuracy could be obtained because of the logarithmic term in the expression for the capacitance of a spherical condenser. It is by no means clear how an auxiliary probe would behave.

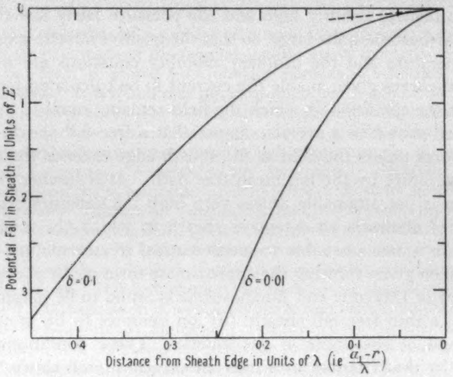


Figure 5. Potential distribution in free-fall sheath.

§6. THE USE OF A SCREENED PROBE

A difficulty which has not so far been mentioned is that of secondary emission from the collector. It has been shown in I that this can be measured by using a probe screened with a fine grid. Successful probes of this type having a small radius (0·05 cm.) have now been built in cylindrical form as well as the plain type used in I, and there seems no reason why a quasi-spherical probe should not be constructed in a similar manner. The thickness of the stem would not be a great disturbance providing the collecting orifice were on the opposite side. A very considerable advantage of such a probe would be the reduction in sheath thickness resulting from its lower operating potential. Owing to its ability to separate positive ions from electrons it could be operated with the electron current to the probe of the same order as the positive ion current instead of negligible compared with it, as is necessary in the case of an unscreened probe.

§7. CONCLUSIONS

The unsuitability of the Langmuir method for determining ion densities at pressures above 1 mm. Hg makes it desirable to examine the possibility of using the positive ion end of the probe current–voltage characteristic for this purpoes.

It is found possible to calculate the potential distribution outside a space-charge sheath if the radius of the sheath is known, and from this to find the current of positive ions to the probe. Both the distribution of potential and the current to the probe are functions of the ion density of the plasma in which the probe is immersed.

If the ion density is fairly high and the pressure fairly low the field around the sheath may become quite large, so that the positive ion energies greatly exceed the gas temperature and the ordinary mobility equations are not valid. The equations and curves given enable the current to be calculated for this condition as well as for the condition in which the field remains small.

It has been shown in a previous paper that a free-fall sheath will not occur at high pressures unless the field at the sheath edge exceeds the quotient of the electron temperature by the ion mean free path. It is further shown here that this condition is not attainable unless very high ion densities exist and that the normal state of affairs is an extensive sheath in which the ions diffuse. The latter condition is not amenable to mathematical treatment, but for the former curves have been given showing the approximate form of the energy distribution.

The theory of Davydov and Zmanovskaja is found to be inapplicable because it assumes (i) a thin free fall sheath, (ii) ion energies to be unmodified by the probe field and of the order of gas kinetic. These conditions are mutually incompatible by the criterion for a free-fall sheath quoted above from I.

The possibility of using a screened probe increases the probable usefulness of the method by (i) reducing the operating sheath thickness, (ii) eliminating secondary emission, and (iii) improving the accuracy in determining the electron temperature. Nevertheless the usefulness of the method is greatly reduced by the absence of any convenient way of determining theoretically or experimentally the extent of the sheath. The only means that shows any promise is the use of a cathetometer, which is inevitably clumsy and rather inaccurate.

It may be given as a general conclusion that at the ion densities normally encountered in glow discharges the method could only be applied after a careful experimental study of the means of measuring sheath thickness. At higher densities the application is much simpler as the sheath thickness can be ignored. At high pressures such densities are not normally met with in regions extensive enough to permit the use of probes, and it may be that a probe could not be made to stand up to the heating and sputtering effect of the bombardment it would receive.

ACKNOWLEDGMENTS

The author wishes to thank Professor H. S. W. Massey for his continued interest and encouragement, Dr. E. H. S. Burhop for discussion and constructive criticism and the Imperial Chemical Industries Fellowships Committee of the University of London for the award of a Fellowship to enable the programme of research, of which this work forms part, to be carried out.

REFERENCES

Boyd, R. L.F., 1950, *Proc. Roy. Soc.* A, **201,** 329.
Davydov, Van B., and Zmanovskaja, L., 1936, *J. Tech. Phys., U.S.S.R.*, **28,** 715.
Sena, L., 1946, *J. Phys., U.S.S.R.*, **10,** 179.

The Temperature of the Upper Atmosphere

By D. R. BATES*
Department of Physics, University College, London

Communicated by H. S. W. Massey; MS. received 15*th February* 1951

ABSTRACT. The properties of a number of simple models of the upper atmosphere are tabulated. A study is made of the thermal equilibrium in the region of the F layers. The rate at which energy is gained from ionizing photons is estimated. Various loss processes are examined. It is found that the most important is conduction, and the next most important is probably emission by the magnetic dipole connecting the two low levels of the ground term of atomic oxygen. These are so effective indeed, that the energy source first considered appears inadequate to maintain the high temperature the upper atmosphere is generally supposed to possess. Alternatives are investigated. Sufficient energy might conceivably be supplied either by band absorption by nitric oxide, or by incoming interstellar matter: but neither source is attractive. Attention is drawn to the possibility that the original estimate of the contribution from ionizing photons may be much too small. This estimate is based on radio measurements. It is shown that these only give a *lower limit* to the electron production rate: for certain ions may be removed so rapidly by recombination that they would escape detection even if their formation rate were greater than that of the dominant ions in the layers. The suggestion is tentatively made that heat is supplied to the upper atmosphere mainly by non-observed ionization. It is pointed out that owing to the possible existence of such ionization, the measurements of the radio scientists cannot be assumed to give the intensity of the solar emission beyond the Lyman limit. The temporal variation of the temperature is briefly discussed.

§ 1. INTRODUCTION

It is generally thought that the temperature of the atmosphere in the region of the F layers and above† is very high. The reasons for this belief have been summarized in various reviews (cf. Spitzer 1949, Bates 1949 a) and will not be given again. All that need be said is that though they are not completely conclusive they are very potent. The present paper is devoted to the examination of the heat gain and heat loss processes, to see whether a high temperature is in fact plausible, and to see in particular whether it is consistent with the currently accepted views on the ionosphere.

§ 2. ATMOSPHERIC MODELS

To provide a basis for discussion it is useful to have available a variety of simple models of the atmosphere. From rocket measurements it seems (cf. Bates and Nicolet 1950) that at the 110 km. level the temperature is about 300° K. and the particle concentration about $7{\cdot}4 \times 10^{12}/\text{cm}^3$. This level was taken as standard, and it was assumed that above it the temperature T first rises linearly, and then remains constant. Making allowance for the diminution of the acceleration of gravity g, and writing $T(z) = T(0) + az$ and $g(z) = g(0)/(1 + bz)^2$ where z is the height above the standard level, and b is the reciprocal of the distance of this level from the centre of the earth, it can be shown that in the linear rise region the total particle concentration, $N(z)$ is given by

---

* Now Professor of Applied Mathematics at Queen's University, Belfast.

† This region is the upper part of what Chapman (1950) calls the *thermosphere*.

$$\log [N(z)/N(0)] = -\log [T(z)/T(0)] - \frac{\bar{m}g(0)}{\boldsymbol{k}}\left\{\frac{a}{\{a-bT(0)\}^2}\right.$$
$$\left.\times [\log \{T(z)/T(0)\} - \log (1+bz)] - \frac{bz}{1+bz}\frac{\log e}{a-bT(0)}\right\}; \quad \ldots\ldots(1)$$

$\bar{m}$ is the mean molecular mass (assumed constant), and $\boldsymbol{k}$ Boltzmann's constant. The corresponding formula in the isothermal region can be derived immediately by putting a equal to zero. Computations were carried out for all combinations of the following temperatures T_B and altitudes h_B of the upper boundary of the linear-rise region: $T_B = 500$, 1,000, 1,500 and 2,000° K., $h_B = 200$, 300, 400 and 500 km. (above ground level). In referring to a particular model the designation $\mathcal{M}(T_B, h_B)$ will be used.

The value of $\bar{m}g(0)/\boldsymbol{k}$ was taken to be 27·1° K/km., which corresponds to an atmosphere of nitrogen and oxygen in the same proportions as at the ground, with the former undissociated and the latter completely dissociated (giving $\bar{m}$ to be 23·8). Table 1 gives the results of the calculations. From it data for a further set of models can be obtained rapidly. Suppose, for example, that diffusive separation is regarded as setting in abruptly at z_D*, and that information on the concentration, $n(X, z)$, of a constituent X of molecular mass m_X is required. Clearly

$$\log [n(X, z)] = \log (r_X) + \log [N(z)]$$
$$+ \frac{\bar{m} - m_X}{\bar{m}} \{\log [N(z_D)] - \log [N(z)] - \log [T(z)/T(z_D)]\}$$

where $r_X = n(X, z_D)/N(z_D)$. It may be noted that if $\bar{m}$ is 23·8 (as used in the computations described) then the values of $(\bar{m} - m_X)/\bar{m}$ for the m_X's appropriate to N, O, N_2, and O_2 are +0·41, +0·33, −0·18 and −0·34, respectively.

Inspection of the figures in the Table reveals that the temperature gradient must be very steep indeed if the particle concentrations usually associated with the F layers are correct. Thus, according to Cowling (1945) the observations made in England on the collision frequency of electrons in the F_1 layer indicate that the logarithm of the particle concentration there (that is at an altitude of some 220 km.) is about 11·1, which is appreciably greater than the value for any of the models. The discrepancy might be partially attributed to the uncertainties in the assumed particle concentration at the standard level (110 km.) and in the collision cross section for atomic oxygen (cf. Bates and Massey 1947), but unless a is more than some 6° K/km. it is very considerable indeed. Again without such a high temperature gradient the atmosphere rapidly becomes so tenuous, that there is difficulty in understanding the formation of the F_2 layer at sometimes as great an altitude as about 400 km. (cf. Appleton 1947) and the occurrence of aurorae up to rather over 1,000 km. (cf. Störmer 1947). It should be borne in mind also that the temperature gradient must actually fall below its peak value for an appreciable altitude range so that a model atmosphere, having a temperature gradient which is constant and equal to the peak, must be rather more extended than the real atmosphere.

---

* In Chapman's terminology z_D is the height of the *turbopause* above the standard level.

Supplementary information on the various models is supplied in Table 2 which gives the total energy (thermal and gravitational *) freed if the temperature

Table 1. Atmospheric Models

Model					Altitude (km.)				
$\mathcal{M}(T_B, h_B)$	ΔT	150	200	250	300	400	500	750	1000
T_B (°K.); h_B (km.)				Log (number of particles/cm³)					
500; 200	2·22	$11{\cdot}4_0$	$10{\cdot}0_0$	$8{\cdot}8_5$	$7{\cdot}7_5$	$5{\cdot}5_5$	$3{\cdot}4_0$	$-1{\cdot}6_5$	$-6{\cdot}3_5$
500; 300	1·05	$11{\cdot}3_5$	$9{\cdot}7_0$	$8{\cdot}3_0$	$7{\cdot}1_0$	$4{\cdot}9_0$	$2{\cdot}7_5$	$-2{\cdot}2_5$	$-6{\cdot}9_5$
500; 400	0·69	$11{\cdot}3_5$	$9{\cdot}6_0$	$8{\cdot}1_0$	$6{\cdot}7_0$	$4{\cdot}2_5$	$2{\cdot}1_5$	$-2{\cdot}9_0$	$-7{\cdot}6_0$
500; 500	0·51	$11{\cdot}3_5$	$9{\cdot}5_5$	$7{\cdot}9_5$	$6{\cdot}4_5$	$3{\cdot}8_5$	$1{\cdot}5_5$	$-3{\cdot}5_0$	$-8{\cdot}2_0$
1000; 200	7·78	$11{\cdot}5_0$	$10{\cdot}5_5$	$10{\cdot}0_0$	$9{\cdot}4_0$	$8{\cdot}3_0$	$7{\cdot}2_5$	$4{\cdot}7_5$	$2{\cdot}4_0$
1000; 300	3·68	$11{\cdot}4_5$	$10{\cdot}2_0$	$9{\cdot}3_0$	$8{\cdot}6_0$	$7{\cdot}5_0$	$6{\cdot}4_5$	$3{\cdot}9_0$	$1{\cdot}5_5$
1000; 400	2·41	$11{\cdot}4_0$	$10{\cdot}0_0$	$8{\cdot}9_5$	$8{\cdot}0_5$	$6{\cdot}6_5$	$5{\cdot}6_0$	$3{\cdot}1_0$	$0{\cdot}7_5$
1000; 500	1·80	$11{\cdot}4_0$	$9{\cdot}9_0$	$8{\cdot}7_0$	$7{\cdot}7_0$	$6{\cdot}0_5$	$4{\cdot}8_0$	$2{\cdot}3_0$	$-0{\cdot}0_5$
1500; 200	13·33	$11{\cdot}5_5$	$10{\cdot}7_5$	$10{\cdot}4_0$	$10{\cdot}0_0$	$9{\cdot}3_0$	$8{\cdot}5_5$	$6{\cdot}9_0$	$5{\cdot}3_0$
1500; 300	6·32	$11{\cdot}4_5$	$10{\cdot}4_5$	$9{\cdot}7_5$	$9{\cdot}2_5$	$8{\cdot}5_0$	$7{\cdot}8_0$	$6{\cdot}1_0$	$4{\cdot}5_5$
1500; 400	4·14	$11{\cdot}4_5$	$10{\cdot}2_5$	$9{\cdot}4_0$	$8{\cdot}7_5$	$7{\cdot}7_5$	$7{\cdot}0_0$	$5{\cdot}3_5$	$3{\cdot}8_0$
1500; 500	3·08	$11{\cdot}4_0$	$10{\cdot}1_0$	$9{\cdot}1_5$	$8{\cdot}3_5$	$7{\cdot}1_5$	$6{\cdot}2_5$	$4{\cdot}6_0$	$3{\cdot}0_0$
2000; 200	18·89	$11{\cdot}5_5$	$10{\cdot}8_5$	$10{\cdot}6_0$	$10{\cdot}3_0$	$9{\cdot}7_5$	$9{\cdot}2_5$	$7{\cdot}9_5$	$6{\cdot}8_0$
2000; 300	8·95	$11{\cdot}5_0$	$10{\cdot}6_0$	$10{\cdot}0_0$	$9{\cdot}6_0$	$9{\cdot}0_5$	$8{\cdot}5_0$	$7{\cdot}2_5$	$6{\cdot}1_0$
2000; 400	5·87	$11{\cdot}4_5$	$10{\cdot}4_0$	$9{\cdot}7_0$	$9{\cdot}1_5$	$8{\cdot}3_5$	$7{\cdot}8_0$	$6{\cdot}5_5$	$5{\cdot}4_0$
2000; 500	4·36	$11{\cdot}4_5$	$10{\cdot}3_0$	$9{\cdot}4_5$	$8{\cdot}8_0$	$7{\cdot}8_5$	$7{\cdot}1_5$	$5{\cdot}8_5$	$4{\cdot}7_0$

ΔT = temperature gradient in linear-rise region in °K/km.

Table 2. Energy Content of Atmosphere above Various Altitudes

Model	Altitude of lower boundary of region (km.)			
$\mathcal{M}(T_B, h_B)$	150	200	300	500
T_B (°K.); h_B (km.)	Log (energy content in ev/cm² column)			
500; 200	16·7	15·5	13·3	9·0
500; 300	16·6	15·1	12·6	8·3
500; 400	16·5	14·9	12·1	7·7
500; 500	16·5	14·8	11·8	7·1
1000; 200	17·3	16·7	15·6	13·4
1000; 300	16·9	16·0	14·7	12·6
1000; 400	16·8	15·6	14·0	11·8
1000; 500	16·7	15·4	13·5	11·0
1500; 200	17·7	17·2	16·5	15·1
1500; 300	17·2	16·6	15·7	14·3
1500; 400	17·0	16·1	15·0	13·5
1500; 500	16·8	15·8	14·4	12·8
2000; 200	17·9	17·6	17·1	16·0
2000; 300	17·4	17·0	16·3	15·3
2000; 400	17·2	16·5	15·7	14·6
2000; 500	17·0	16·2	15·1	13·9

above certain selected levels were reduced to zero and the gas condensed to a thin layer located at these levels. In compiling the Table use was made of the

---

* It may be remarked that the contribution from the thermal energy is about three times that from gravitational energy.

approximate formulae that the mean temperature of the atmosphere above the level z_E is

$$\left.\begin{aligned}\overline{T}(z_E) &= T(z_E) + H(z_E)\frac{a}{1-c}\left\{1-\frac{n(z_B)}{n(z_E)}\left(\frac{H(z_B)}{H(z_E)}\right)^2\right\} && (z_B > z_E)\\ &= T(z_E) && (z_B < z_E)\end{aligned}\right\} \qquad \ldots\ldots(2)$$

and that the distance of the centre of gravity from z_E is

$$\left.\begin{aligned}\overline{z}_E &= H(z_E)\frac{1}{1-c}\left\{1-c\frac{n(z_B)}{n(z_E)}\left(\frac{H(z_B)}{H(z_E)}\right)^2\right\} && (z_B > z_E)\\ &= H(z_E) && (z_B < z_E)\end{aligned}\right\} \qquad \ldots\ldots(3)$$

where $\qquad H(z) = \boldsymbol{k}T(z)/\overline{m}g(z) \qquad$ and $\qquad c = aH(z)/T(z). \qquad \ldots\ldots(4)$

The gas was presumed to consist of N_2 and O in the same proportions as before and the specific heat at constant pressure was taken to be $2{\cdot}7 \times 10^{-4}$ ev/deg. K/particle. Vibrational and electronic energies were neglected.

§ 2. HEAT GAIN BY PHOTO-IONIZATION

The incident flux of solar photons required to produce a Chapman layer having a peak electron concentration of $n(\mathrm{e})$ at zero zenith angle is $S = eH\alpha n(\mathrm{e})^2$ where e is the Naperian base, H is the local scale height and α is the effective recombination coefficient (cf. Mitra 1947). This well-known formula can readily be modified (Mohler 1940) so as to apply to an atmosphere in which T has a constant gradient a, and in which α is proportional to the pth power of N, the gas concentration. It becomes

$$S = \exp\{(1+c)(1-p)\}H\alpha n(\mathrm{e})^2/(1+c)(1-p) \qquad \ldots\ldots(5)$$

where c is as defined in (4), and the variable parameters have the values pertaining to the layer maximum. Sometimes H is not itself measured but instead is derived from the observed half thickness τ of a parabolic layer fitted to the actual layer. Provided the Chapman representation is valid these two quantities approximately satisfy the simple relationship $H = \frac{1}{2}\tau$ (cf. Mitra 1947). However for the conditions specified above it is necessary to replace this by $H = \frac{1}{2}\tau(1+c)(1-p)$. Substitution in (5) gives immediately that

$$S = \tfrac{1}{2}\exp\{(1+c)(1-p)\}\tau\alpha n(\mathrm{e})^2. \qquad \ldots\ldots(6)$$

The magnitudes to be assigned to the parameters appearing in (5) and (6) must now be considered.

Radio investigations give $n(\mathrm{e})$ directly. Allen (1948) has made a careful survey of the available data and concludes that at sunspot minimum $n(\mathrm{e}|\mathrm{F}_1)$ is $2{\cdot}4 \times 10^5/\mathrm{cm}^3$, and $n(\mathrm{e}|\mathrm{F}_2)$ is $5{\cdot}9 \times 10^5/\mathrm{cm}^3$. Unfortunately the other parameters are much less well determined.

Most of the attempts to measure $\alpha(\mathrm{F}_1)$ have been based on observations of the decay of the ionization during an eclipse. The following are amongst the values reported: $8 \times 10^{-9}\,\mathrm{cm}^3/\mathrm{sec}$. (Higgs 1942), $6 \times 10^{-9}\,\mathrm{cm}^3/\mathrm{sec}$. (Rydbeck 1946), $7 \times 10^{-9}\,\mathrm{cm}^3/\mathrm{sec}$. (Alpert and Einberg 1947), $2 \times 10^{-9}\,\mathrm{cm}^3/\mathrm{sec}$. (Denisse, Seligmann and Gallet 1947) and $1{\cdot}4 \times 10^{-8}\,\mathrm{cm}^3/\mathrm{sec}$. (McLeish 1948). Their mean is $7 \times 10^{-9}\,\mathrm{cm}^3/\mathrm{sec}$.

The eclipse method has also been used to determine $\alpha(F_2)$, and in addition several workers have sought to derive this coefficient from the asymmetry in the diurnal variation of the ionization, or from the rate of replenishment of the ionization after its disappearance in a disturbance. Table 3 shows the results obtained. The value of $n(e|F_2)$ on the day of each measurement is included, as it probably influences the apparent $\alpha(F_2)$, and is thus of relevance. The true altitude of the layer would also have been of interest; but in many cases information on it is lacking. It can be seen that there is a considerable scatter amongst the coefficients quoted. As far as can be judged from them $\alpha(F_2)$ cannot be much above 3×10^{-10} cm³/sec. at sunspot minimum.

Table 3. Observed Recombination Coefficients in the F_2 Layer

Method used	n(e)	$\alpha(F_2)$	Reference
Eclipse	17	0·4	Higgs (1942)
	5·5	2 to 5	Wells & Shapley (1946)
	3·1	5 to 10	
	2·5	10	
	1·9	10	McLeish (1948)
Asymmetry	10	0·4	Bhar (1939)
	10	0·8	Appleton (1937)
	18	0·9	Mohler (1940)
	15	0·9	Allen (1948)
	13	1·6	
	4·1	2·3	
	4·5	3·1	
	27	10	Denisse, Seligmann & Gallet (1947)
Recovery	11	1·7	Berkner & Seaton (1940)

n(e) in units of 10^5/cm³; $\alpha(F_2)$ in units of 10^{-10} cm³/sec.

From an analysis of the data on the diurnal variation of the altitude of the F_1 layer at Huancayo Kellogg (1950) claims that $H(F_1)$ is 43 km. However the results of the investigations on the F_2 layer (see below) seem to indicate that this is an over-estimate, and a value of 35 km. will be adopted.

Measurements on the magnitude of $\tau(F_2)$ have been made by a number of radio scientists. Table 4 summarizes the principal published data. Inspection of it suggests that $\tau(F_2)$ is unlikely to exceed about 100 km. at sunspot minimum (when $h_M(F_2)$, the altitude of the layer, is comparatively low). It may be remarked, incidentally, that owing to the probable dependence of $\alpha(F_2)$ on N the very large values of $\tau(F_2)$ found do not necessarily imply extreme temperatures. For example if p were 0·6 (as might well be the case), and if c were small, $\tau(F_2)$ would be about $5H(F_2)$ so that if it were 180 km., $H(F_2)$ would only be 36 km., and hence (neglecting the possible partial dissociation of N_2) $T(F_2)$ would only be about 900° K.

Table 4

Half-thickness $\tau(F_2)$ (km.)	51	56	47	90	115	110	200
Altitude of layer $h_M(F_2)$ (km.)	220	220	230	240	330	360	380
Reference	Pekeris (1940)			Appleton (1947)		Booker & Seaton (1940)	

Owing to the form of (5) $S(F_1)$ is insensitive to c and p provided they are not large. In the case of the F_2 layer (6) must be used. The quantity $(1+c)(1-p)$ was taken to be 0·5, which choice, though rather arbitrary, cannot lead to serious error.

On inserting the figures that have been given in formulae (5) and (6) it can be seen that the fluxes required to produce the F_1 and F_2 layers at sunspot minimum are: ${}_0S(F_1)=4\times10^9/\mathrm{cm}^2/\mathrm{sec.}$ and ${}_0S(F_2)=9\times10^8/\mathrm{cm}^2/\mathrm{sec.}$ According to Allen (1948) the fluxes at sunspot number R are related to those at sunspot number zero, by the equations

$${}_RS(F_1)=(1+0{\cdot}0124R)\,{}_0S(F_1) \quad \text{and} \quad {}_RS(F_2)=(1+0{\cdot}0200R)\,{}_0S(F_2).$$

These approximations do not take full account of such factors as the possible variation of the scale height through the sunspot cycle. They are however the best available. At sunspot maximum R is 100, so that at this period the fluxes apparently are ${}_{100}S(F_1)=9\times10^9/\mathrm{cm}^2/\mathrm{sec.}$ and ${}_{100}S(F_2)=3\times10^9/\mathrm{cm}^2/\mathrm{sec.}$

Some of the ionization in the F_2 layer *must* be produced by the radiation responsible for the F_1 layer. As can easily be seen the fraction of the total ionization rate at the maximum of the layer, contributed by this radiation is

$$f=\{\alpha(F_1)n(e|F_1)^2/\alpha(F_2)n(e|F_2)^2\}r\exp(1-r) \qquad \ldots\ldots(7)$$

where r is here the ratio of the gas density in the F_2 layer to that in the F_1 layer (Bates 1949 b). Numerical substitution yields $f=10r\exp(-r)$ at sunspot minimum, and $f=8r\exp(-r)$ at sunspot maximum. Thus if at the two periods mentioned r were, for instance, 0·1 and 0·04, f would be about 1·0 and 0·3 respectively. It would seem therefore that the total flux ${}_0S(F_1+F_2)$ should be taken to be about $4\times10^9/\mathrm{cm}^2/\mathrm{sec.}$ and the total flux ${}_{100}S(F_1+F_2)$ to be about $1{\cdot}1\times10^{10}\,\mathrm{cm}^2/\mathrm{sec.}$

The parent neutral particles from which the F layers are formed have not yet been identified with certainty (cf. Bates and Seaton 1950). They may be O or N atoms or possibly N_2 molecules. The ionization potentials of these are 13·6 ev., 14·5 ev. and 15·6 ev. respectively, but in photo-ionization by a body at such a high temperature as the Sun, the mean energy absorbed is probably about 0·5 ev. above the threshold (Eddington 1930). All this energy is not available for local heating. For example, if the ultimate recombination process* (Bates and Massey 1947) were of the type

$$XY^+ + e \rightarrow X' + Y' \qquad \ldots\ldots(8)$$

some of it would be used in exciting the fragments X and Y, and some in producing the dissociation. The first part would be emitted as radiation. It might be suggested that the second part is not lost but reappears on association following a three-body collision $X+Y+M\rightarrow XY+M$. This however is not the case. The rate coefficient for such a reaction is unlikely to exceed $10^{-32}\,\mathrm{cm}^6/\mathrm{sec.}$, so that even if the concentrations of Y and M were as high as $3\times10^{11}/\mathrm{cm}^3$ in the region of the F layers, a particular X particle would have to wait on an average 10^9 sec. before recombining there. This time is so long that a far more probable fate is downward diffusion (Spitzer 1949). Radiative association $X+Y \rightarrow XY+\boldsymbol{h}\nu$ may be more rapid; but it is not a very effective mechanism for converting the energy of formation of a molecule into heat.

---

* It should be noted that it is by no means certain that the main cause of the disappearance of the free charges in the F_2 layer is always recombination, and is not instead diffusion. The only recombination process that seems at all plausible is (8) which requires molecular ions. Bates and Massey (1947) have suggested that these are formed from atomic ions by charge transfer collisions. However because of the low gas density at the 300 km. and above it is difficult to find a reaction which can replenish the XY molecules involved sufficiently rapidly.

From the considerations given it seems not unreasonable to suppose that the local heat gain does not much exceed about 6 ev/photon and may well be smaller. Hence, using the estimates made of the total flux and averaging over a full day it can be seen that the thermal energy made available at the equator is apparently at most some 1×10^{10} ev/cm^2/sec. during sunspot minimum and 3×10^{10} ev/cm^2/sec. during sunspot maximum. At the temperate latitudes of 50° N. or S. these mean values must be multiplied by 0·7 for winter conditions, and by 1·5 for summer conditions.

On referring to Table 2 it will be noted that if the temperature is high the total energy content of the upper atmosphere is so great that the response to the heating action caused by the ionizing photons is very slow.

§ 3. HEAT LOSS BY RADIATION

3.1. *Atoms*

Under upper atmospheric conditions appreciable energy cannot be emitted through transitions having a high excitation potential. Consequently it has hitherto been supposed that O, $^1D-{}^3P$, is the only one that need be taken into account. The first detailed calculations (Spitzer 1949) suggested that this transition might be of considerable importance. They depended, however, upon the results of a quantal study (Yamanouchi, Inui and Amemiya 1940) of the rate of excitation by electron impact, and these results are now known to be grossly in error since they exceed by several powers of ten a limit set by a certain conservation theorem (Bates, Fundaminsky, Leech and Massey 1950). On making the necessary simple alteration to the original calculations it is immediately apparent that the cooling effect of O, $^1D-{}^3P$, is actually negligible.

The possible importance of transitions between the different levels of the ground term of atomic oxygen appears to have been overlooked in earlier investigations. In fact, however, these transitions have properties which are likely to make them very effective. Thus the $^3P_0$ and $^3P_1$ levels of O lie only 0·028 ev. and 0·020 ev. respectively above the $^3P_2$ level (cf. Moore 1949) so that the excitation potentials involved are of the same order as the energy of thermal motion. Further the Einstein A coefficients are quite high, that of the transition O, $^3P_0-{}^3P_1$ being $1{\cdot}7 \times 10^{-5}$ sec., and that of the transition O, $^3P_1-{}^3P_2$ being $8{\cdot}9 \times 10^{-5}$/sec. (Pasternack 1940). It is therefore necessary to determine how much energy is lost through this far infra-red emission. Because of the relative magnitudes of the Einstein coefficients* A_{0-1} and A_{1-2}, of the statistical weights ω_0 and ω_1, and of the excitation energies ϵ_0 and ϵ_1, the first transition is insignificant compared with the second and will be ignored.

The distribution of the oxygen atoms between the two lower levels will be that characteristic of the kinetic temperature T if excitation

$$\mathrm{O}(^3\mathrm{P}_2)+\mathrm{M} \rightarrow \mathrm{O}(^3\mathrm{P}_1)+\mathrm{M}-0{\cdot}020\ \mathrm{ev.}\ (\mathrm{M}=\mathrm{N}_2, \mathrm{O}, \mathrm{etc.}) \quad \ldots\ldots(9)$$

takes place much faster than does photon emission, that is if

$$\eta P_{2-1} \exp\left(-\epsilon_1/\boldsymbol{k}T\right) n(\mathrm{O}, {}^3\mathrm{P}_2)\, n(\mathrm{M}) \gg A_{1-2}\, n(\mathrm{O}, {}^3\mathrm{P}_1), \quad \ldots\ldots(10)$$

---

* The identifying subscripts 0, 1 and 2 on the various symbols refer of course to the J values of the different levels of the ground ^3P term.

η being the gas kinetic collision coefficient, and $P_{2\text{-}1}$ being the probability that (9) occurs when there is a collision in which the requisite energy is available. For the specified distribution, $n(\mathrm{O}, {}^3\mathrm{P}_1) = (\omega_1/\omega_2) \exp(-\epsilon_1/\boldsymbol{k}T) n(\mathrm{O}, {}^3\mathrm{P}_2)$ so that (10) is equivalent to

$$P_{2\text{-}1} \gg \frac{A_{1\text{-}2}\omega_1/\omega_2}{\eta n(\mathrm{M})}. \qquad \text{......(11)}$$

Now ω_1 and ω_2 are 3 and 5 respectively, η depends on the temperature but for present purposes can be taken to be about $5 \times 10^{-10}\,\mathrm{cm}^3/\mathrm{sec.}$, and $A_{1\text{-}2}$ is as given above. Hence (11) becomes $P_{2\text{-}1} \gg 1 \times 10^5/n(\mathrm{M})$. This condition is not very stringent. For it to be satisfied in a region where $n(\mathrm{M})$ is as low as $1 \times 10^9/\mathrm{cm}^3$ (and on the high temperature theory such a gas density is not encountered until at least the 250 km. level is reached), requires only that $P_{2\text{-}1}$ be much greater than 1×10^{-4}.

No direct information on (9) is available but a number of similar processes have been studied in the laboratory (cf. Mitchell and Zemansky 1934). They have been found to proceed quite rapidly. For example,

$$\mathrm{Hg}(6\,{}^3\mathrm{P}_0) + \mathrm{N}_2 \rightarrow \mathrm{Hg}(6\,{}^3\mathrm{P}_1) + \mathrm{N}_2' - 0{\cdot}218\,\mathrm{ev}. \qquad \text{......(12)}$$

occurs with a probability of 3×10^{-3}, and

$$\mathrm{Na}(3\,{}^2\mathrm{P}_{1/2}) + \mathrm{A} \rightarrow \mathrm{Na}(3\,{}^2\mathrm{P}_{3/2}) + \mathrm{A} - 0{\cdot}002\,\mathrm{ev}. \qquad \text{......(13)}$$

occurs with a probability which is within an order of unity. Noting the small energy discrepancy in (9) it seems not unreasonable to suppose that $P_{2\text{-}1}$ does indeed exceed 1×10^{-4} by a considerable factor.*

The high temperature region of the atmosphere emits and absorbs the O, ${}^3\mathrm{P}_1 - {}^3\mathrm{P}_2$ photons. If it is optically thin, those emitted are insufficient to make an appreciable contribution to the intensity of the radiation field, which therefore may be taken as arising entirely from the outward flux from the relatively cool region below. In these circumstances the net rate at which energy is lost is

$$R^{\mathrm{E}}(\mathrm{O}) = \epsilon_1 A_{1\text{-}2} n(\mathrm{O}, {}^3\mathrm{P}_1)\{1 - \sigma/\rho\} \qquad \text{......(14)}$$

where σ is the energy of that part of the radiation field which lies within the Doppler width of the line, and ρ is the energy which would lie within this width if the radiation field were in thermodynamic equilibrium with the gas. The intensity of the flux from the lower region is half the intensity of black body radiation at the temperature T_{L} of some optically deep level. Hence if the local temperature of the gas in the upper region is T the ratio σ/ρ is less than $T_{\mathrm{L}}/2T$. As this is a small number (14) may without serious error be written

$$\begin{aligned}
R^{\mathrm{E}}(\mathrm{O}) &= \epsilon_1 A_{1\text{-}2} n(\mathrm{O}, {}^3\mathrm{P}_1) \\
&= \epsilon_1 A_{1\text{-}2} n(\mathrm{O}, {}^3\mathrm{P}_{2,1,0}) \left\{ \frac{\omega_1 \exp(-\epsilon_1/\boldsymbol{k}T)}{\omega_2 + \omega_1 \exp(-\epsilon_1/\boldsymbol{k}T) + \omega_0 \exp(-\epsilon_0/\boldsymbol{k}T)} \right\} \\
&= 5{\cdot}8 \times 10^{-7} n(\mathrm{O}, {}^3\mathrm{P}_{2,1,0}) \\
&\quad \times \left[\frac{3\omega_1 \exp(-\epsilon_1/\boldsymbol{k}T)}{\omega_2 + \omega_1 \exp(-\epsilon_1/\boldsymbol{k}T) + \omega_0 \exp(-\epsilon_0/\boldsymbol{k}T)} \right] \mathrm{ev}/\mathrm{cm}^3/\mathrm{sec.}
\end{aligned}$$

In this last formula the term in square brackets is of order unity.

---

* It is perhaps of relevance to note that both ${}^3\mathrm{P}_2$ and ${}^3\mathrm{P}_1$ can tend to ${}^3\Sigma_0$ under the influence of an electric field (such as would be provided by a nearby atom or molecule). This suggests that there is quite a high chance of a transference from one to the other level during a collision.

If the high temperature region is of an appreciable optical thickness t the treatment is more complex. Suppose however that both this region and the low temperature region are isothermal, the temperature changing discontinuously at the boundary between them. The relevant radiative transfer equation is of a simple form. Its solution shows that for such an atmosphere the original expression for the rate of escape of energy should be multiplied by $\{1-(1-t)\exp(-t)+t^2\mathrm{Ei}(-t)\}/2t$, where $\mathrm{Ei}(-t)$ is the exponential integral. When t is 0·1, 0·2 and 0·4 this factor equals 0·84, 0·74 and 0·61, respectively. The corresponding corrections for other model atmospheres cannot be very different, and unless t is large they are clearly in any case of only minor importance.

If λ is the central wavelength and $\Delta\nu$ is the Doppler frequency width of the oxygen line, then α_{2-1} the absorption cross section in its core is

$$\frac{1}{4\pi}\left(\frac{\ln 2}{\pi}\right)^{1/2}\frac{\omega_1}{\omega_2}\frac{\lambda^2 A_{1-2}}{\Delta\nu}, \qquad \ldots\ldots(15)$$

(cf. Mitchell and Zemansky 1934). Numerical substitution yields

$$\alpha_{2-1}=9{\cdot}_4\times 10^{-17}/T^{1/2}\,\mathrm{cm}^2.$$

Taking the important stimulated emission into account it can be seen that the optical thickness of the region of the atmosphere above the level of altitude h_R is given by

$$t=\int_{h_R}^{\infty}\alpha_{2-1}\{1-\exp(-\epsilon_1/\boldsymbol{k}T)\}n(\mathrm{O},{}^3\mathrm{P}_2)\,dh.$$

Instead of evaluating the integral exactly it was judged sufficient to take for T the mean temperature. Calculations carried out on this basis show that the atmosphere above the 200 km. level has a negligible optical thickness, and that even the atmosphere above the 150 km. level has an optical thickness of only 0·2 or less. The evaluation of the rate at which energy is radiated is therefore simple. Table 5 displays the results obtained for the various atmospheric models already described. It will be noted that if the temperature is high the estimated rate of gain of energy due to the processes responsible for the formation of the F layers (cf. §2) is appreciably less than the rate of loss from above the 200 km. level (at least at sunspot minimum), and is much less than the rate of loss from above the 150 km. level.

Table 5. Energy Loss due to the Transition O, $^3\mathrm{P}_1-{}^3\mathrm{P}_2$ in regions above given level

Model $\mathcal{M}(T_\mathrm{B}, h_\mathrm{B})$	Rate of energy loss log (ev/cm² column/sec.)		Model $\mathcal{M}(T_\mathrm{B}, h_\mathrm{B})$	Rate of energy loss log (ev/cm² column/sec.)	
T_B (°K.); h_B (km.)	150 km.	200 km.	T_B (°K.); h_B (km.)	150 km.	200 km.
500; 200	10·4	9·3	1500; 200	11·2	10·7
500; 300	10·3	8·9	1500; 300	10·9	10·2
500; 400	10·2	8·7	1500; 400	10·7	9·9
500; 500	10·1	8·6	1500; 500	10·5	9·6
1000; 200	10·9	10·3	2000; 200	11·4	11·0
1000; 300	10·6	9·7	2000; 300	11·1	10·6
1000; 400	10·5	9·4	2000; 400	10·8	10·2
1000; 500	10·4	9·2	2000; 500	10·7	9·9

3.2. *Molecules*

Polyatomic molecules are extremely rare at great altitudes owing to photo-dissociation (cf. Bates and Nicolet 1950, Bates and Witherspoon 1951) and in consequence emission from them can be neglected. But emission from diatomic molecules merits some discussion. If there is a Boltzmann distribution amongst the vibrational levels of a constituent X, the rate of radiation of energy in the 1→0 transition (which alone need be taken into account) is simply

$$R^{\mathrm{V}}(\mathrm{X}) = f(\mathrm{X}) N \epsilon^{\mathrm{V}}{}_1 \exp\left(-\epsilon^{\mathrm{V}}{}_1/\boldsymbol{k}T\right) A^{\mathrm{V}}{}_{1-0} \qquad \ldots\ldots (16)$$

where $f(\mathrm{X})$ is the fractional concentration of the constituent, N as usual is the total particle concentration, $\epsilon^{\mathrm{V}}{}_1$ is the excitation energy and $A^{\mathrm{V}}{}_{1-0}$ is the Einstein coefficient. In the case of nitrogen the first vibrational level lies 0·29 ev. above the zeroth (cf. Herzberg 1950) and $f(\mathrm{N}_2)$ is 0·64 (in the absence of dissociation) so that when the temperature is, for example, 1500° K., (16) becomes $R^{\mathrm{V}}(\mathrm{N}_2) \simeq 2 \times 10^{-2} N A^{\mathrm{V}}{}_{1-0}$ ev/cm³/sec. giving $R^{\mathrm{V}}(\mathrm{N}_2)/R^{\mathrm{E}}(\mathrm{O}) \simeq 1 \times 10^5 A^{\mathrm{V}}{}_{1-0}$. The value of $A^{\mathrm{V}}{}_{1-0}$ is unknown but it is certainly very low as a quadrupole transition is involved. Noting that James and Coolidge (1938) find that the Einstein coefficient for the corresponding transition in hydrogen is only $3{\cdot}5 \times 10^{-7}$/sec., it would seem most unlikely that $R^{\mathrm{V}}(\mathrm{N}_2)/R^{\mathrm{E}}(\mathrm{O})$ is as great as unity.* With the molecule $^{14}\mathrm{N}\,^{15}\mathrm{N}$ a dipole transition can of course occur. It must however be much too feeble to compensate for the fact that the $^{15}\mathrm{N}$ isotope is some $1{\cdot}4 \times 10^{-3}$ times less abundant than the $^{14}\mathrm{N}$ isotope. Thus the difference in the masses of the nuclei is relatively small, and even for HD (in which the mass difference is relatively large), the dipole moment is so weak (Wick 1935, Herzberg 1938) that the transition probability associated with it is only about 2×10^2 times that associated with the quadrupole moment.

Since oxygen is highly dissociated in the region of the F layers (Penndorf 1949), it is apparent without detailed discussion that $R^{\mathrm{V}}(\mathrm{O}_2)/R^{\mathrm{E}}(\mathrm{O})$ is very small.

The Einstein coefficient for a vibrational transition in a molecule composed of atoms of different nuclear charge is usually quite high. Consequently it is less easy to dismiss the possibility of significant emission from such substances as carbon monoxide and nitric oxide. The most useful approach is through the consideration of the rate of population of the first vibrational level by inelastic collisions. This rate can be expressed in the form $f(\mathrm{X}) N^2 K \exp\left(-\epsilon^{\mathrm{V}}{}_1/\boldsymbol{k}T\right)$ where $K \exp\left(-\epsilon^{\mathrm{V}}{}_1/\boldsymbol{k}T\right)$ is the coefficient associated with the excitation mechanism, X is CO or NO, and the other symbols have their customary meaning. The emission rate, $R^{\mathrm{V}}(\mathrm{X})$, can be taken to be the same. For CO $\epsilon^{\mathrm{V}}{}_1$ is 0·27 ev. and for NO $\epsilon^{\mathrm{V}}{}_1$ is 0·23 ev. (cf. Herzberg 1950). When the temperature is 1500° K. it may hence be seen that for either substance $R^{\mathrm{V}}(\mathrm{X})/R^{\mathrm{E}}(\mathrm{O}) \simeq 2 \times 10^5 f(\mathrm{X}) N K$. At the 150 km. level N is about 3×10^{11}/cm³ (cf. Table 1). As is well known it is very difficult to convert translational energy into vibrational energy so that K must be extremely low (cf. Massey 1949). A value of 10^{-14} cm³/sec. (which corresponds to an efficiency of about 2×10^{-5}/collision) may be adopted without risk of serious underestimation. With these figures

$$R^{\mathrm{V}}(\mathrm{X})/R^{\mathrm{E}}(\mathrm{O}) \simeq 6 \times 10^2 f(\mathrm{X}). \qquad \ldots\ldots (17)$$

---

* The loss due to conduction generally exceeds that due to the oxygen magnetic dipole (cf. § 4) so that even if $R^{\mathrm{V}}(\mathrm{N}_2)/R^{\mathrm{E}}(\mathrm{O})$, and other such ratios, were actually rather greater than unity the processes concerned would still not be important.

The carbon monoxide in the upper atmosphere results from the photo-dissociation of carbon dioxide. Near ground level $f(CO_2)$ is only 3×10^{-4} (Paneth 1937). As $f(CO)$ can scarcely exceed this, $R^V(CO)/R^E(O)$ must be small. In the absence of evidence to the contrary it seems best to assume that nitric oxide remains a very minor atmospheric constituent at all altitudes. It would thus appear likely from (17) that $R^V(NO)/R^E(O)$ is also less than unity.

Rotational transitions of nitrogen and oxygen are obviously even less important than vibrational transitions. Those in heteronuclear molecules must be considered with more care since rotational excitation (in contrast to vibrational excitation) is readily effected by collisions (cf. Massey 1949). The square of the transition matrix element involved is simply $J\mathscr{P}^2$, where J is the rotational quantum number, and $\mathscr{P}$ is the dipole moment. Hence if B is the rotational constant of the molecule the Einstein coefficient equals $2^9\pi^4(BJ)^3J\mathscr{P}^2/3\boldsymbol{h}(2J+1)$, $\boldsymbol{h}$ being here Planck's constant. Assuming a Boltzmann distribution and averaging over all values of J it can be seen that the approximate mean rate of emission of energy is given by

$$R^R(X)=\frac{2^{10}\pi^4B^2\boldsymbol{k}^2T^2}{3\boldsymbol{h}^2\boldsymbol{c}}\mathscr{P}^2N,$$

or more conveniently $R^R(X)=0{\cdot}26(\boldsymbol{k}T)^2B^2\mathscr{P}^2f(X)N$ ev/cm³/sec. where B is in cm^{-1} and $\mathscr{P}$ is in atomic units. For CO the numerical value of B is 1·93 (cf. Herzberg 1950), and that of $\mathscr{P}$ is about 5×10^{-2} (cf. Fowler 1935), so that at 1500° K. $R^R(CO)\simeq4\times10^{-5}f(CO)N$ and therefore $R^R(CO)/R^E(O)\simeq2\times10^2f(CO)$ which is clearly negligible. Provided its abundance is low, nitric oxide can similarly be shown to emit only a small amount of energy.

It is concluded that the ${}^3P_1-{}^3P_2$ magnetic dipole transition of atomic oxygen is probably responsible for the main radiative cooling of the upper atmosphere in the region of the F layers.

§ 4. HEAT LOSS BY CONDUCTION

In a stimulating recent paper Spitzer (1949) has drawn attention to the importance of thermal conduction in the upper atmosphere. Fortunately its influence can be estimated with fair reliability. The conductivity coefficient of a gas is given by the formula

$$\kappa=0{\cdot}0448(\boldsymbol{k}mT)^{1/2}c_v(9\gamma-5)/\sigma^2 \qquad \ldots\ldots(18)$$

where c_v is the specific heat at constant volume, γ is the ratio of the specific heats and σ is the effective collision diameter (cf. Chapman and Cowling 1939). Consider the conductivities of atomic and molecular oxygen. Apart from σ the parameters appearing in (18) are known. Substitution yields $\kappa(O)=1{\cdot}1\{\sigma(O_2)/\sigma(O)\}^2\kappa(O_2)$ and the corresponding relation for nitrogen has the same numerical factor. Hence as the collision diameters for the atoms can scarcely be larger than for the molecules (and are probably indeed appreciably smaller) it seems safe to conclude that κ is not reduced by any dissociation in the upper atmosphere.

Table 6 gives the conductivities of ordinary air at various temperatures obtained by the extrapolation of standard laboratory data (as quoted by Chapman and Cowling) by means of Sutherland's formula.

Table 6. Conductivity of Air at Various Temperatures

Temperature T (° K.)	500	1000	1500	2000
Conductivity κ (ev/(° K. km^{-1})/cm²/sec. $\times10^{10}$)	2·4	3·7	4·7	5·5

If the temperature gradient dT/dz is expressed in deg. K/km. the energy flow downwards is simply $\kappa dT/dz$ ev/cm² sec. It is immediately apparent that conduction must exercise a marked influence on the upper atmosphere. For example, if T is even 500°, the flow is so great that equating it to the estimated mean energy input in the equatorial region gives that dT/dz at the base of the F layers is at most about 0·4° K/km. during sunspot minimum, and 1·2 K/km. during sunspot maximum. These gradients (especially since they must decrease with increasing altitude) are clearly inconsistent with the high temperatures usually advocated. Moreover, in calculating them the effect of radiative cooling was ignored.

§ 5. EQUILIBRIUM

Information relevant to the thermal equilibrium of the upper atmosphere is collected in Table 7, which for each of the models gives (i) the rates of energy loss from the regions above the 150 km. and 200 km. levels, due to the combined effect of radiation (§ 3) and conduction (§ 4), (ii) the characteristic times obtained by dividing the total energy content (Table 2) by these rates, (iii) the ratios of these rates to the rates of energy gain from F layer ionization (§ 2) at sunspot minimum and maximum.

Table 7. Combined Energy Loss from Radiation and Conductivity, and its comparison with total energy content and apparent energy gain from F-layer ionization.

Model $\mathcal{M}(T_B, h_B)$ T_B (°K.); h_B (km.)	(1)		(2)		(3) Sunspot min.		Sunspot max.	
	(*a*)	(*b*)	(*a*)	(*b*)	(*a*)	(*b*)	(*a*)	(*b*)
500; 200	10·8	10·7	5·9	4·8	0·8	0·7	0·4	0·3
500; 300	10·6	10·4	6·0	4·8	0·6	0·4	0·1	−0·1
500; 400	10·4	10·1	6·1	4·8	0·4	0·1	0·0	−0·3
500; 500	10·4	10·0	6·1	4·8	0·4	0·0	−0·1	−0·5
1000; 200	11·5	11·5	5·8	5·2	1·5	1·5	1·0	1·0
1000; 300	11·1	11·0	5·8	5·0	1·1	1·0	0·6	0·5
1000; 400	10·9	10·8	5·9	4·8	0·9	0·8	0·4	0·3
1000; 500	10·8	10·6	5·9	4·8	0·8	0·6	0·3	0·2
1500; 200	11·8	11·8	5·9	5·4	1·8	1·8	1·3	1·3
1500; 300	11·4	11·4	5·8	5·2	1·4	1·4	0·9	0·9
1500; 400	11·2	11·1	5·8	5·0	1·2	1·1	0·7	0·6
1500; 500	11·0	10·9	5·8	4·9	1·0	0·9	0·5	0·4
2000; 200	12·0	12·1	5·9	5·5	2·0	2·1	1·5	1·6
2000; 300	11·6	11·6	5·9	5·4	1·6	1·6	1·1	1·1
2000; 400	11·4	11·3	5·8	5·2	1·4	1·3	0·9	0·9
2000; 500	11·2	11·1	5·8	5·0	1·2	1·1	0·7	0·7

(1) Rate of energy loss R (log (ev/cm² column/sec.)); (2) Characteristic time parameter (log (total energy content/R)); (3) Log (R/rate of gain of energy from F-layer ionization).

Note: log (12 hours in seconds) $= 4{\cdot}6_4$. (*a*) $h_E = 150$ km.; (*b*) $h_E = 200$ km.

Inspection of the figures shows that the characteristic times are in general of the order of a few days so that pronounced cooling in the dark hours is not to be expected.

The ratios introduced in (iii) can conveniently be denoted by ${}_Z\mathscr{R}[h_E, \mathscr{M}(T_B, h_B)]$, where h_E is the altitude (150 km. or 200 km.) of the lower boundary of the atmospheric region considered, Z is the sunspot number (0 at minimum and 100 at maximum), and $\mathscr{M}(T_B, h_B)$ indicates the model used (§2). As can be seen they approximate to unity only when the temperature gradient assumed is low, and in other circumstances they can be very large indeed: for example ${}_0\mathscr{R}$ [150 or 200 km., $\mathscr{M}$ (1,500° K., 200 km.)] is about 60, and ${}_{100}\mathscr{R}$ [150 or 200 km., $\mathscr{M}$ (1,500° K., 200 km.)] is about 20. Now the cooling processes discussed *must* be occurring, and care has been taken not to over-estimate them. Hence if the view that the temperature increases rapidly above the E layer is to be preserved, some means must be found for augmenting the heating processes.

5.1. *Non-Ionizing Radiation*

The rate at which a constituent X absorbs energy through a transition of oscillator strength f and wave number σ is simply

$$I = n(\mathrm{X}) \left\{ \frac{8\pi^2 \boldsymbol{e}^2 \boldsymbol{h} \sigma^3 \Omega f}{\boldsymbol{m}} \right\} \exp(-\boldsymbol{hc}\sigma/\boldsymbol{k}T_{\mathrm{R}}), \qquad \ldots\ldots(19)$$

where T_{R} and Ω are the temperature and dilution of the radiation field and $\boldsymbol{e}$ and $\boldsymbol{m}$ are the electronic charge and mass respectively. The energy appearing as heat may of course be much less than I because of emission from the excited molecules formed. To make most direct use of laboratory experience it is convenient to write

$$f = \frac{\boldsymbol{mc}^2 V \sigma^2}{\pi \boldsymbol{e}^2 \boldsymbol{N}} \frac{W}{l}; \qquad \ldots\ldots(20)$$

W is the equivalent width of the absorption band formed in traversing a path length l at S.T.P., $\boldsymbol{c}$ the velocity of light, V the molar volume, and $\boldsymbol{N}$ Avogadro's number (cf. Mitchell and Zemansky 1934). The introduction of (20) in (19) yields

$$I = n(\mathrm{X}) \left\{ \frac{8\pi \boldsymbol{hc}^2 V \sigma^5 \Omega W}{\boldsymbol{N} l} \right\} \exp(-\boldsymbol{hc}\sigma/\boldsymbol{k}T_{\mathrm{R}}). \qquad \ldots\ldots(21)$$

It is believed that T_{R} is about 5,000° K. for radiation below the Lyman limit* (cf. Bates and Nicolet 1950); Ω is $5{\cdot}4 \times 10^{-6}$. Hence (21) becomes

$$I = 2 \times 10^{-(25 + 1{\cdot}25 \times 10^{-4}\sigma)} \sigma^5 \frac{W}{l} n(\mathrm{X}) \text{ ev/cm}^3\text{/sec}. \qquad \ldots\ldots(22)$$

where σ is expressed in cm^{-1}, W in A., and l in cm.

Nitrogen is probably the most abundant molecular constituent in the region of the F layers. At the 200 km. level its concentration, $n(\mathrm{N_2})$, is unlikely to exceed 4×10^{10}/cm^3 (cf. Table 1). The lowest transition is the $\mathrm{X}\,{}^1\Sigma_g^+ - \mathrm{A}\,{}^3\Sigma_u^+$ of Vegard–Kaplan, for which σ is about 5×10^4 cm^{-1}. It follows from (22) that the contribution to I is at most some $1 \times 10^3 W/l$ ev/cm^3/sec. Even if W/l were as great as unity this would only be of the same order as the rate of gain of energy from F-layer ionization; and in fact W/l must be very small since the Vegard–Kaplan system has never been detected in absorption. Nitrogen is indeed almost completely transparent up to 7×10^4 cm^{-1} (Chapman and Price 1936). Between this wave number and 1×10^5 cm^{-1} narrow bands have been found but they also are too weak (Hopfield 1922) to give any considerable heating.* If the 5,000° K. model of the Sun is approximately correct, the incident flux in the remaining spectral range below the Lyman limit is insufficient for the intense absorption that sets in to give a significant contribution. The calculations of Penndorf (1949) show that oxygen is very highly dissociated at the altitudes concerned. In spite of the great Schumann–Runge system it is therefore not important. Carbon monoxide is clearly uninfluential since it lacks strong and suitably located transitions (Leifson 1926), and since in addition it is a rare constituent (§3.2). The position regarding nitric oxide is rather less certain. A path length of 0·15 cm. is sufficient to produce absorption bands in the 5×10^4 cm^{-1} region whose total equivalent width is perhaps as much as 100 A. (Leifson 1926). Using (22) it can hence be seen that the energy absorbed near the 200 km. level is of order $10^6 f(\mathrm{NO})$ ev/cm^3/sec. where as usual $f(\mathrm{NO})$ is the fractional concentration. However partial re-emission through electronic, followed by vibrational, transitions ensues. Since both these are rapid it is doubtful if more than $10^4 f(\mathrm{NO})$ ev/cm^3/sec. is retained locally. This amount of energy is negligible if $f(\mathrm{NO})$ is small, as is probably the case. But it would be appreciable if nitric oxide were much more abundant than is commonly supposed; and such a possibility, though remote, cannot be altogether excluded.

---

* Above the Lyman limit the contribution from non-ionizing radiation cannot be significantly greater than that from ionizing radiation.

5.2. *Particles*

Watson (1941) has given estimates of the mass of material entering the atmosphere in the form of meteors of various visual magnitudes down to the limit set by the Poynting–Robertson effect. It is an elementary task to show that negligible heating is produced above the 150 km. level even if these estimates (which were partially obtained by the extrapolation of observational data) are too low by a factor 10^4, as Whipple (private communication) believes may be possible.

A more promising source of energy is the interstellar matter moving towards the Sun under its gravitational field. Ryle (1950) has recently reported certain irregularities in the F layers which he suggests are caused by such matter. The ionization involved is only about 10^{-3} of the total. However the fraction of the available kinetic energy used in ionizing is probably far less than the fraction used in heating; it may indeed be 10^4 times smaller (cf. Herlofson 1948). Hence there is at least the possibility that the thermal effect of the incoming stream is sufficient to keep the temperature of the upper atmosphere high; but of course more evidence on the properties of the stream is required before this can be regarded as other than speculative.

5.3. *Unobserved Ionization*

As has been seen it is difficult, though not impossible, to envisage a mechanism which supplies more heat to the upper atmosphere than does the ionizing radiation. In view of this it is necessary to re-examine this last source carefully to ensure that its potentiality has not been underestimated.

The greatest uncertainty lies in the recombination coefficients. While it is conceivable that the values deduced by the radio scientists are rather low it seems unlikely that they are in error by a factor of between one and two powers of ten which is the extent of the discrepancy found (cf. Table 7). However the methods used in measuring the recombination coefficients do not give complete information; for they take no account of the fact that a layer may contain several kinds of ion. This may have important consequences.

---

* I wish to thank Dr. W. C. Price for referring me to the experimental data on nitrogen absorption and in particular for pointing out that Leifson's results (1926) are incorrect.

Suppose that ions X_1^+ and X_2^+ are produced at rates q_1 and q_2, and that α_1 and α_2 are the recombination coefficients associated with them. In equilibrium their concentrations n_1 and n_2 are related by the equation $\alpha_1 n_1/q_1 = \alpha_2 n_2/q_2$. Hence if $\alpha_1 \gg \alpha_2 q_1/q_2$ then $n_1 \ll n_2$ even though $q_1 \gg q_2$. In these circumstances the dominant X_2^+ ions would control the main properties of the layer and it would not be apparent to radio scientists that X_1^+ ions were being formed and destroyed in large numbers. Consider for example the determination of the apparent recombination coefficient* from the diurnal variation. The asymmetry about noon, upon which the method is based, is associated with the X_2^+ ions, for $\alpha_1 n_1 n(\mathrm{e})$ must closely follow q_1 since α_1 is large. Again the X_1^+ ions exert little influence on the decay during an eclipse or the recovery after a disturbance as their concentration is much smaller than that of the X_2^+ ions. It is clear therefore that the measurements of the radio scientists give only a *lower limit* to the actual rate of ion production.

Recombination coefficients differ widely amongst themselves and it is not at all unlikely that some constituent of the atmosphere forms ions which disappear rapidly. These ions might perhaps be N_2^+, and the mechanism by which they are removed might be dissociative recombination,

$$N_2^+ + e \rightarrow N' + N''. \qquad \ldots\ldots(23)$$

Quantal arguments have recently been given (Bates 1950) which indicate that such a process may be extremely effective, a coefficient as high as 10^{-7} cm^3/sec. being possible. Laboratory measurements (Biondi and Brown 1949) lend some support to this view. The tentative proposal made is not of course inconsistent with the evidence from the spectrum of the dawn flash, which shows (Bates 1949 b) that the concentration of N_2^+ ions in the F layers is low. It may be remarked that if it is correct, reaction (23) may be an important source of nitrogen atoms.

By postulating the occurrence of *unobserved ionization* it is obviously possible to account for the apparent discrepancy associated with the high temperature theory of the upper atmosphere since such ionization might well supply sufficient energy to balance that being lost.† Though this postulate is of necessity arbitrary it is in no way implausible. Indeed it has several minor attractive features. Thus according to it the main agency heating the upper atmosphere is the solar radiation beyond the Lyman limit. The intensity of this varies through the sunspot cycle (cf. Appleton and Naismith 1939, Appleton 1939) and hence so also should the temperature and thus the altitudes of the layers. Radio observations (Appleton 1947) suggest that the predicted variation does in fact take place. Again on the old interpretation of the ionospheric data it appeared that the incident flux of photons of energy 13·5 ev. and above was but little greater than would occur if the Sun were a 6,000° K. black body. A number of theorists (cf. Woolley

---

* It may be noted that the *apparent* recombination coefficient is not in general the same as the *average* recombination coefficient, which is merely the parameter α appearing in the equation, $q_1 + q_2 = \alpha n(\mathrm{e})^2$, and which may be shown (Bates and Massey 1947) to equal

$$\alpha_2 n_2 (q_1 + q_2)/q_2 n(\mathrm{e}) \quad \text{or} \quad \alpha_1 n_1 (q_1 + q_2)/q_1 n(\mathrm{e}).$$

† If the temperature of the upper atmosphere were known reliably Table 7 could be used to estimate the amount of unobserved ionization required (or vice versa).

and Allen 1948, Giovanelli 1949, Hoyle 1949, Thomas 1949) have attempted to calculate the intensity of the solar emission in the spectral region under consideration.[*] The predicted intensities are unfortunately in rather poor agreement with each other; but it is interesting to note that in general they are much higher than can be reconciled with the measurements on the F layers if unobservable ionization is not invoked.

REFERENCES

Alpert, J. L., and Einberg, A. A., 1947, *Bull. Acad. Sci., U.R.S.S., Ser. Geogr. Geophys.*, **11,** 137.

Allen, C. W., 1948, *Terr. Magn. Atmos. Elect.*, **53,** 433.

Appleton, E. V., 1937, *Proc. Roy. Soc.* A, **162,** 451; 1939, *Occ. Not. Roy. Astr. Soc.*, **3,** 33; 1947, *J. Instn. Elect. Engrs.*, Pt. 3A, **94,** 186.

Appleton, E. V., and Naismith, R., 1939, *Phil. Mag.*, **27,** 144.

Bates, D. R., 1949 a, *Mon. Not. Roy. Astr. Soc.*, **109,** 215; 1949 b, *Proc. Roy. Soc.* A, **196,** 562; 1950, *Phys. Rev.*, **78,** 492.

Bates, D. R., Fundaminsky, A., Leech, J. W., and Massey, H. S. W., 1950, *Phil. Trans. Roy. Soc.* A, **243,** 117.

Bates, D. R., and Massey, H. S. W., 1947, *Proc. Roy. Soc.* A, **188,** 350.

Bates, D. R., and Nicolet, M., 1950, *J. Geophys. Res.*, **55,** 301.

Bates, D. R., and Seaton, M. J., 1950, *Proc. Phys. Soc.* B, **63,** 129.

Bates, D. R., and Witherspoon, Alice E., 1951, *Mon. Not. Roy. Astr. Soc.*, in the press.

Berkner, L. V., and Seaton, S. L., 1940, *Terr. Magn. Atmos. Elect.*, **45,** 393.

Bhar, J. N., 1939, *Indian J. Phys.*, **22,** 253.

Biondi, M. A., and Brown, C., 1949, *Phys. Rev.*, **76,** 1697.

Booker, H. G., and Seaton, S. L., 1940, *Phys. Rev.*, **57,** 87.

Chapman, S., 1950, *J. Atmos. Terr. Physics*, **1,** 121.

Chapman, S., and Cowling, T. G., 1939, *Mathematical Theory of Non-Uniform Gases* (Cambridge: University Press).

Chapman, S., and Price, W. C., 1936, *Rep. Prog. Phys.*, **3,** 55 (London: Physical Society).

Cowling, T. G., 1945, *Proc. Roy. Soc.* A, **183,** 453.

Denisse, F., Seligmann, P., and Gallet, R., 1947, *C.R. Acad. Sci., Paris*, **225,** 1169.

Eddington, A. S., 1930, *Internal Constitution of the Stars* (Cambridge: University Press).

Fowler, R. H., 1935, *Statistical Mechanics*, Second Edition (Cambridge: University Press).

Giovanelli, R. G., 1949, *Mon. Not. Roy. Astr. Soc.*, **109,** 298.

Herlofson, N., 1948, *Rep. Prog. Phys.*, **11,** 444 (London: Physical Society).

Herzberg, G., 1938, *Astrophys. J.*, **87,** 428; 1950, *Molecular Spectra and Molecular Structure—Diatomic Molecules*, Second Edition (New York: Van Nostrand).

Higgs, A. J., 1942, *Mon. Not. Roy. Astr. Soc.*, **102,** 42.

Hopfield, J. J., 1922, *Phys. Rev.*, **20,** 573.

Hoyle, F., 1949, *Some Recent Researches in Solar Physics* (Cambridge: University Press).

James, H. M., and Coolidge, A. S., 1938, *Astrophys. J.*, **87,** 438.

Kellogg, W. W., 1950, *Survey of Data and Theoretical Analysis of the Upper Atmosphere*, Pt. 4, Ed. J. Kaplan (Los Angeles: Institute of Geophysics, University of California).

Leifson, S. W., 1926, *Astrophys. J.*, **63,** 73.

McLeish, C. W., 1948, *Canad. J. Res.*, **26A,** 137.

Massey, H. S. W., 1949, *Rep. Prog. Phys.*, **12,** 248 (London: Physical Society).

Matheson, L., 1932, *Phys. Rev.*, **40,** 813.

Mitchell, A. C. G., and Zemansky, M. W., 1934, *Resonance Radiation and Excited Atoms* (Cambridge: University Press).

Mohler, F. L., 1940, *Bur. Stand. J. Res., Wash.*, **25,** 507.

Mitra, S. K., 1947, *The Upper Atmosphere* (Calcutta: Royal Asiatic Society of Bengal).

Moore, Charlotte E. 1949, *Atomic Energy Levels* (Washington: National Bureau of Standards).

Paneth, F. A., 1937, *Quart. J. Roy. Met. Soc.*, **63,** 433.

Pasternack, S., 1940, *Astrophys. J.*, **92,** 129.

---

[*] **My thanks are due to Professor J. L. Greenstein for helpful discussions on this work.**

Penndorf, R., 1949, *J. Geophys. Res.*, **54,** 7.
Pekeris, C. L., 1940, *Terr. Magn. Atmos. Elect.*, **45,** 205.
Rydbeck, O. E. H., 1946, *K. Tekn. Högsk. Handl.*, No. 53.
Ryle, M., 1950, *Observatory*, **70,** 136.
Spitzer, L., 1949, *Atmospheres of the Earth and Planets*, Ed. G. P. Kuiper (Chicago : University Press), p. 213.
Störmer, C., 1947, *Observatory*, **67,** 161.
Thomas, R. N., 1949, *Astrophys. J.*, **109,** 489.
Watson, F. G., 1941, *Between the Planets* (Philadelphia : Blakison).
Wells, H. W., and Shapley, A. H., 1946, *Terr. Magn. Atmos. Elect.*, **51,** 401.
Wick, G. C., 1935, *Atti Accad. Naz. Lincei*, **21,** 708.
Woolley, R. d. v. R., and Allen, G. W., 1948, *Mon. Not. Roy. Astr. Soc.*, **108,** 292.
Yamanouchi, T., Inui, T., and Amemiya, A., 1940, *Proc. Phys.-Math. Soc., Japan*, **22,** 847.

LETTERS TO THE EDITOR

Photoconductive Cells of Cadmium Selenide

In a previous letter (Schwarz 1950) the spectral distribution of sensitivity of polycrystalline layers of CdSe and some values of sensitivity to white light have been published. Since then an appreciable improvement in the performance of these cells has been achieved and some more properties have been measured.

An improved cell with a sensitive area of 9·8 mm. × 0·185 mm. has been measured at room temperature by the National Physical Laboratory with the results shown in the Tables below.

Table 1. Sensitivity to Monochromatic Radiation and Spectral Sensitivity

P.D. across cell (v.)	Absolute sensitivity (amp/w.) for : $\lambda=0{\cdot}5461\ \mu$	$\lambda=0{\cdot}70\ \mu$	$\lambda=0{\cdot}90\ \mu$
6	1690	5580	80
24	4090	*	*

* Not measured.

Table 2. Variation of Sensitivity with Intensity of Radiation for $\lambda=0{\cdot}5461\ \mu$ and a p.d. of 6 volts

Radiation intensity (μw/cm^2)	0·01	0·1	1·0	6·55	10	100
Sensitivity (amp/w.)	220	220	720	1700	1700	1100

Table 3. Sensitivity to White Light (colour temperature of 2848° K.)

	Illumination (lumens/m^2)	0·215	2·15	21·5	215
P.D. across cell = 6 v.	current (μa.)	0·86	58	680	3340
	sensitivity (amp/lumen)	2·2	15	17·5	8·6
P.D. across cell = 24 v.	current (μa.)	9·2	254	1740	7800
	sensitivity (amp/lumen)	23·6	65·5	44·7	20

The dark resistance of this cell at room temperature for a potential difference of 6 volts and 24 volts is $1{\cdot}7\times10^9$ ohms and $8{\cdot}2\times10^8$ ohms respectively.

As can be seen from the tables, the photocurrent depends in a complicated way on light intensity and applied voltage, and is linear only within certain limits. This dependence can be varied for different cells.

Some preliminary a.c. measurements with a chopping frequency of 800 c/s. indicate that the minimum detectable energy at room temperature for a bandwidth of 1 c/s. is of the order of 2×10^{-12} watts for $\lambda=0{\cdot}7\mu$.

Besides possessing very high sensitivity, these cells have the following properties : (i) the range of sensitivity extends from the x-ray region into the infra-red to approximately 1·4 μ, (ii) very small overall size and size of sensitive area, (iii) a potential difference of a few volts is sufficient, (iv) the permissible current through the cell without undue fatigue is of the order of 20 to 30 milliamperes.

The development of these cells has been guided by a suggested theory of photoconductivity (Schwarz 1949), and the achievement of the sensitivity quoted above indicates that this theory is essentially correct.

The author would like to thank Telecommunications Research Establishment, Great Malvern, for the A.C. measurements, and Hilger and Watts Ltd. for permission to publish this letter.

Hilger and Watts Ltd. (Hilger Division),
Camden Road, London, N.W.1.
19th June 1951.

E. SCHWARZ.

SCHWARZ, E., 1949, *Proc. Phys. Soc.* A, **62,** 530 ; 1950, *Proc. Phys. Soc.* B, **63,** 624.

Crystal Diode and Triode Action in Lead Selenide

Since the discovery of transistor action in Ge (Bardeen and Brattain 1948), it has been of interest to search for other materials which show the effect. In a recent publication describing transistor action in p-type lead sulphide (Banbury, Gebbie and Hogarth 1951), it was pointed out that the many similarities in electrical properties between PbS, PbSe and PbTe, would indicate that the last two substances would also be of interest as possible transistor materials.

Single crystal specimens of PbSe, prepared by W. D. Lawson (to be published) of Telecommunications Research Establishment, were examined. Specimens of both n- and p-type were used. They possessed the following D.C. properties at room temperature :

	n-type PbSe	p-type PbSe
Hall coefficient R (cm^3/coulomb)	-4	$+1{\cdot}4$
Conductivity σ (ohm^{-1} cm^{-1})	300	600
Mobility product $R\sigma$ (cm^2 volt^{-1} sec^{-1})	1200	840
Approximate carrier concentration (cm^{-3})	10^{18}	3×10^{18}

When examined with a fine spot of white light and a tungsten whisker contact, photovoltaic effects were observed and were greater for the p-type specimens. The n-type PbSe showed moderate rectifying properties against a whisker contact (Figure 1) with a turn-over at 3 or 4 volts in the reverse direction, whereas the p-type material acted as an excellent crystal detector, and exhibited voltage amplification effects when examined as a crystal triode in a manner similar to that described for PbS by Gebbie, Banbury and Hogarth (1950). The rectification phenomena were examined with whisker contacts of tungsten and phosphor bronze, substantially similar properties being exhibited with each material.

A voltage–current characteristic for p-type PbSe is shown in Figure 2. In the case of this p-type material, forming processes were of great importance for the production of good rectifying contacts, and were similar to those already described for PbS (Banbury, Gebbie

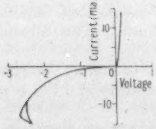


Figure 1.

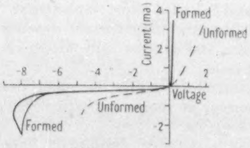


Figure 2.

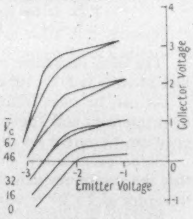


Figure 3.

Figure 1. Voltage–current characteristic for tungsten whisker contacts on n-type lead selenide.
Figure 2. Voltage–current characteristic for tungsten whisker contacts on p-type lead selenide.
Figure 3. Collector voltage–emitter voltage characteristics for a lead selenide crystal triode. (Collector load, 100 kΩ; maximum voltage gain on curves is 4·4.)

and Hogarth 1951). Forming in the reverse direction could be accomplished by the application of a voltage signal, while forming in the forward direction was effected by a sharp increase in current caused by a sudden reduction in series resistance. The peak back voltages were normally in the range 4 to 7 volts, although occasionally values as high as 12 volts were obtained. Rectification ratios for small signal voltages as high as 300 : 1 have been recorded and values of 50 : 1 could be obtained without difficulty. The best rectifying contacts were obtained on cleavage planes. Polishing the surfaces impaired the rectification but the properties could be restored by etching with a warm 60% solution of perchloric acid, or by heating to 150° C. *in vacuo* for an hour or so.

When examined for transistor action, the crystals showed voltage amplification, and small power gains (of order 1·5 to 2) could be obtained. Values of current gain α greater than unity were not observed, the usual values being of order 0·3. It was frequently found (in contrast to highly pure Ge and PbS), that an emitter bias of 3 to 4 volts was necessary to produce amplification. The highest value of voltage amplification so far recorded is 7. Typical curves relating collector voltage to emitter voltage are shown in Figure 3. The spacing between point contacts in the above experiments was approximately 0·01 cm.

The PbSe transistors here described are seen to be inferior to those based on Ge and PbS. In view of the comparatively high impurity content of the present specimens, this cannot be taken as a final assessment of the ultimate merits of PbSe in this respect. Experiments on specimens of high purity seem desirable, although none is at present available. Meanwhile, the fact that transistor action can be obtained in a wider range of substances than previously reported is thought to be of some importance in the development of transistor theories, inasmuch as it reduces the likelihood of theories being based on *ad hoc* assumptions concerning a particular material.

The author would like to express his appreciation to the Chief Superintendent, Telecommunications Research Establishment, for the supply of lead selenide specimens, to Professor R. W. Ditchburn for the provision of laboratory facilities, and to his colleagues in the semiconductor group for helpful advice and discussions.

Physics Department,
The University, Reading.
7th June 1951.

C. A. HOGARTH.

BANBURY, P. C., GEBBIE, H. A., and HOGARTH, C. A., 1951, *Semi-Conducting Materials*, Ed. H. K. Henisch (London : Butterworth's Scientific Publications) p. 78.
BARDEEN, J., and BRATTAIN, W. H., 1948, *Phys. Rev.*, **74**, 230.
GEBBIE, H. A., BANBURY, P. C., and HOGARTH, C. A., 1950, *Proc. Phys. Soc.* B, **63**, 371.

REVIEWS OF BOOKS

The Origin of the Earth, by W. M. SMART. Pp. viii + 239. (Cambridge : University Press, 1951.) 12*s*. 6*d*.

This book, by the Regius Professor of Astronomy in the University of Glasgow, grew out of talks to Service men in the recent war, and it is intended for the general reader of scientific interests.

The work is wider in scope than the title would suggest, inasmuch as the formation of the Earth is involved with that of the solar system as a whole. In fact, the first of the three parts into which Professor Smart divides his discourse is mainly devoted to deducing a common origin for the planets from the observed uniformities of their motions, with the Sun, or possibly some other star, as the parent body. Part II deals with the problem of ascertaining 'the age of the Earth'. It recalls how the conflict between Lord Kelvin and the Victorian geologists over this question was resolved by the discovery of radioactivity, and it explains how divers lines of investigation converge in estimating at some 2,000 million years the time that has elapsed since the formation of the Earth's crust. An elementary discussion of nuclear physics serves to introduce the reader to the carbon–nitrogen cycle, by which the Sun's output of radiation is believed to be maintained.

In the third and concluding section the writer sketches the principal theories (from those of Kant and Laplace to that of Mr. Hoyle) which have been advanced to account for the birth of the solar system; and he indicates the various considerations which have prevented any one of them from carrying complete or lasting conviction. Of particular interest is the brief account of Alfvén's recent hypothesis; this assigns the decisive role in the formation of the planets not to gravitational but to electromagnetic forces, conceived as acting between the rotating magnetic Sun and a surrounding cloud of ions formed from the atoms of a nebula through which the Sun was passing at that critical epoch of its history.

In a brief Epilogue Professor Smart contrasts the older theories of celestial origins, based predominantly upon classical mechanics, with the recent speculations conditioned by the flux of contemporary physical ideas. To both these phases of its enthralling theme the book affords an excellent introduction.

There appears to be a slip on page 60, where it is stated, of ice, that " an increase in pressure leads to the raising of the melting-point ". A. ARMITAGE.

The Classical Theory of Electricity and Magnetism, by MAX ABRAHAM, revised by RICHARD BECKER. Pp. viii + 289. 2nd English Edition. (London: Blackie, 1950.) 25*s*.

Why do authors still keep writing books on electromagnetic theory when the best book has already been written? Their only excuse is that Abraham's and Becker's classic, in the form in which it has come to the English-reading public is only a student's textbook. If Becker could have spared the time to revise Abraham's second volume, and to bring up to date his own sequel to it, the *Elektronentheorie* (which appeared in German in 1933 and has not yet found an English translator) this, alas! only imaginary tripartite work could well hold the same position in electromagnetism as Rayleigh's *Theory of Sound* holds in acoustics.

As it is, we must be grateful that we have at least the best honours student's textbook on the electromagnetic theory of bodies at rest. It is the work of two generations, and two generations of physicists have already grown up on it. August Föppl started it in 1894, when Maxwell's theory, only recently re-stated by Heaviside and Hertz, was still a great novelty. The work was continued by Max Abraham, a physicist not unlike Heaviside in character, but very unlike in his attitude to mathematical proofs. Finally, in 1930, it was modernized in a masterful way by Richard Becker, who carefully left everything standing which was already perfect, but with equal care eliminated what was dated, and added to it, in a last chapter, the first correct thermodynamical theory of field energy and numerous exercises which turned it into a student's textbook. The translator, Dr. John Dougall, revised these in the first, and now again in the second English edition, in which, without increasing their number (137), many exercises have been replaced by problems taken from English examination papers.

In the controversial question of units and dimensions Becker still takes his stand four-square on the Gaussian system. His arguments in the Preface cannot be bettered. One can only add that no amount of perfection in the Giorgi system could make up for the harm which it has done by the fact that at least for some time to come books and papers will be written in two different systems, and by widening the rift between engineers and physicists.

D. GABOR.

Wave Motion and Sound, by R. W. B. STEPHENS and A. E. BATE. Pp. vi + 448. (London: Edward Arnold, 1950.) 45*s*.

This book aims to be a text intermediate between the elementary textbook, " one of strings and pipes, and the advanced treatise, with copious references to original researches ". The main text provides an adequate course for the pass degree student; thirty appendices, mainly mathematical, are included to cover honours course requirements. There are many excellent figures; simplification in some, notably of the ear, materially assists in their understanding. A number of examples are included.

The arrangement chosen for the material has led to a number of anomalies in presentation, e.g., one finds the details on ' Vibrating Bodies ' scattered throughout the first thirteen chapters. Also the space devoted to the differentiation of sine and cosine and to the measurement of definition of elastic moduli seems unnecessary. Again notation varies, e.g., c, C and v for velocity and f and n for frequency—a small point, but consistency particularly helps the beginner.

Of necessity some subjects have had to be compressed and have unfortunately suffered in accuracy. One may instance mean tone and equal temperament mixed on page 211 and some misunderstanding on the question of the various components in bell notes.

Some references to papers are given, others just indicated, e.g., page 423 and page 425. Although the authors say they have deliberately merely quoted lists of suitable books, an introduction to references and their use seems an essential ingredient of an honours course. The short historical introduction seems rather biased towards workers in the U.S.A. although it is well appreciated that much work has been done in that country.

But these are mainly minor comments; the book may be safely reccommended to students. H. D. P.

Electromagnetic Problems of Microwave Theory, by H. MOTZ. Pp. vii + 184. (London: Methuen; New York: Wiley and Sons, 1951.) 9*s*. 6*d*.

This little book is the latest addition to the famous Methuen series of monographs on physical subjects. The author states in the preface that his aim is to illustrate the methods of analysis used in microwave theory by thoroughly worked examples, selected in such a way that the whole forms a reasonably balanced account of the field. One might therefore have expected an outline of the methods of analysis in the introductory chapter. Instead, the author has described a few of the elements used in microwave circuits, such as klystrons, travelling wave tubes, the cavity magnetron, wavemeters, and has included an elementary account of resonator theory and transmission line concepts. The introduction lacks an integrating section in which the author might unfold his plan for presenting a balanced account of the field of microwave theory.

In Chapters 2 and 3 the author presents a concise and clearly written account of the principles of velocity modulation, which includes a discussion of the debunching effect of space charge. Chapter 4 is a useful summary of the main results of the theory of the cavity magnetron, as worked out by D. R. Hartree and O. Bunemann during the war. This includes a derivation of the threshold criterion and a qualitative discussion of the instability voltage. The argument leading to the last equation on page 67 is not clear. This equation and the following deduction from it are true only when the electrons rotate with the Larmor frequency. This chapter is a useful introduction to the more detailed analysis of the cavity magnetron, given by L. R. Walker in his book *Microwave Magnetrons*.

A short statement of the standard mode theory of waveguides is given in Chapter 5. It is however surprising that no reference is made to Schelkunoff's excellent treatise on the subject in *Electromagnetic Waves*.

Chapter 6 deals with the calculation of fields in cavities and waveguides by relaxation methods. This is the author's special field of study, and this chapter succeeds perhaps best in all the book in achieving the declared aim of illustrating a method of analysis by a thoroughly worked example. The method of calculating the phase velocity of a TM-wave in a corrugated waveguide, which has been used by Walkinshaw in the design of a linear accelerator, is also described.

A method of calculating the impedance of a wire antenna in a rectangular waveguide is given in Chapter 7. Use is made of images, and the current density is assumed to be sinusoidal along the wire. In the reviewer's opinion a simpler and more elegant treatment can be given using the delta-function technique. By this method the field of a dipole in a guide can be written down immediately in terms of the modes in the guide. It is worth while too to point out that the real problem of the antenna in the waveguide is to find the current distribution, and that the correct expression for A' is critically dependent on the method of feeding the wire. A factor $\tan^2 \frac{1}{2}kl$ is missing from the first term of $-A'$ in (VII, 43) and the denominator following the summation sign should be $1-(\pi n/bk)^2$.

The final chapter is concerned with the theory of discontinuities in waveguides. Bethe's theory for small holes is well presented together with the integral equation method. A brief statement of the variational method is given, but the author does not point out that in some cases this method can be used to find not only a lower bound for admittance Y_{ii} but also an upper bound.

In such a small monograph the author was clearly obliged to select only a few topics in the wide field of microwaves. One might have expected to find a brief statement of Babinet's principle. It is to be hoped that a place will be found for it in a future edition. Nevertheless the author has kept faith with the title, if not with the preface, of his book in giving us a useful collection of electromagnetic problems of microwave theory. G. G. MACFARLANE.

Bibliography of Electron Microscopy, edited for the Institute of Physics by V. E. COSSLETT. Pp. 350. (London: Edward Arnold, 1950.) 40*s*.

This is a publication in book form of the card index bibliography of the electron microscopy group of the Institute of Physics. The range of subjects covered by this bibliography is a very wide one. Apart from the obvious papers on electron optics and on the design of electron microscopes, papers on all kinds of possible application of the electron microscope are included.

The wide divergence of the contents of these papers may be illustrated by the following titles, chosen at random from the more peripheral subjects: Extension of Langmuir's tables for plane diodes with a Maxwellian distribution of the electrons; High speed microtome for electron microscopy; Making small apertures in metal plates; Contribution of virus research to chemistry and physics; On the existence, morphology, nature and functions of the cytoplasmic membrane in the bacterial cell; Thermally produced oxide films on aluminium; The fusion of carbon, etc.

German and French references on electron microscopy are very complete, and a fair amount of Russian and other foriegn literature is quoted, all titles being given in English.

The entries are listed alphabetically according to the name of the author. Considering, however, the great inhomogeneity of the subject, it seems to the reviewer that the content of this bibliography would be more accessible if the entries were separated into a few major sections, e.g. (i) electron optics, (ii) design and construction of electron microscopes, (iii) investigation of biological objects, (iv) investigation of inorganic materials, etc.

The present work apparently satisfies a need which has been expressed by various institutions and research laboratories, many of which have actively contributed to the preparation of the bibliography. The work has been compiled under the expert direction of the editor, V. E. Cosslett. It certainly represents a useful guide to the relevant literature which should find its place in every major scientific library. Print, paper and binding are of high standard; however, the relatively high price is likely to prevent its purchase by many individual workers.

O. KLEMPERER.

Physics in Chemical Industry, by R. C. L. BOSWORTH. Pp. xix + 928. (London: Macmillan, 1950.) 70*s*.

The author of a book of this kind is faced with greater difficulty in deciding upon the scope and level of treatment than the writer of a closely-knit monograph on a special scientific topic for experts. In a book on a diversity of topics for those who are not experts, the choice of subject matter should, however, be governed by an underlying theme, and the treatment should be suitable for the class of reader which the author has in mind.

The aims and ideas of the author of this book in these respects are clearly stated. He is convinced that ". . . . a deeper knowledge of physics is a necessity to the scientific worker in industry, whether he calls himself chemist, engineer, or physicist". He writes mainly for the student of chemistry intending to enter chemical industry, and for the practising chemical engineer who is handicapped by insufficient knowledge of physics. He is, moreover, anxious to establish an understanding between the academic and the industrial scientist by showing how industrial problems may be solved by the application of basic physical principles, and that such problems are often worthy of serious attention from a scientific point of view. It is stressed that the industrial scientist is seldom able to choose his experimental conditions, and that his success depends on his ability to deal with non-idealized situations.

The contents of the book indicate how the author has attempted to fulfil these aims. It is divided into four parts. The first, consisting of seven chapters, is the mathematical introduction. After a preliminary discussion of standards and measurement, there are chapters on vector analysis, mechanical computation, graphical and numerical methods, dimensional analysis and statistics. In the second part (ten chapters) on the properties of matter, an elementary account of atomic theory, excluding wave mechanics, is followed by a rather more advanced treatment of thermodynamics, which leads naturally to statistical mechanics and the properties of gases. Other topics which follow are the theory of the solid state, the properties of metals, the properties of liquids, surface phenomena, elasticity and rheology. The third part of the book, which is classified as dealing with the movement of matter, contains ten chapters on wave motion, diffusion, fluid flow, transport phenomena, heat flow, diffusion of electric charge, radiation and the rate of chemical reactions. In the fourth part, there are nine chapters on scientific instruments.

When the scope is as wide as this, to achieve anything like completeness would require terse literary style and concise mathematical argument. Ease of presentation would be sacrificed in order to attain depth of treatment, and in physics as in other activities, the deeper a problem, the greater the interest. The author has not pursued this course, but he has nevertheless succeeded in writing an interesting book, and one which may be read with profit by the student embarking on a career in chemical industry or by the chemical engineer who must encounter a variety of physical problems in the course of his work. It is not really a reference book in which the reader may expect to find something about everything that he wants to know. Many topics which a physicist would regard as important are either omitted or only lightly touched upon. It is doubtful whether a scientist whose interests are mainly academic would learn much about the problems that beset his friends in industry by reading this book. On the other hand, the chemist or the industrial research worker would learn a good deal of physics, and what is more important, would be stimulated to look further into many of the questions that are discussed. M. R. HOPKINS.

Glass : A Handbook for Students and Technicians, edited by J. HOME DICKSON. Pp. 300. (London : Hutchinson's Scientific and Technical Publications, 1951.) 25*s*.

This book, edited by J. Home Dickson, and written by him and seven collaborators, is, as its title indicates, intended for the general scientific reader. It covers, in four parts, 1. General Principles, 2. Manufacture, 3. Applications, 4. Appendix (spectral transmission of coloured glasses, the modern field of knowledge of glass). The authors are sufficiently authoritative in their respective fields to ensure that the contents are both accurate and up to date. The range covered is wide, dealing with such topics as the strength of windows in aeroplanes to withstand the impact of birds, the transmission of solar heat through glass, the Bicheroux process for rolling plate glass, glass-to-metal seals, fibreglass, and chemical engineering applications, to mention a few at random. The co-operation of manufacturers in the provision of data and photographs has much increased the value of the book, both as regards reliability and interest. It is easy in such a case as the present to point to omissions due to the endeavour to view a whole industry in one volume, and that industry one which in its technical journal during the last thirty years has published hundreds of papers and still is conscious of its own lack of knowledge. But to do so would be ungracious, as its authors admit the difficulty of selection and are fully aware of the danger. In spite of this comment, however, the book can be recommended as giving a broad picture of a fascinating subject, and the balance between the scientific and practical aspects has been well maintained. It will help all those who are students or technicians, whether in the glass industry or not. Those whose lives are spent with glass will find much of interest and value, and those who normally work in other fields can accept what they read here as reliable and may also gain fresh ideas for their own work as a result of these studies of glass. The publication of such a book should finally dispose of the idea, once justified, that the glass industry is empirical and secretive. Today the reverse is certainly nearer the truth. The book contains selected references to the literature for those who wish to study the subject in more detail. W. M. HAMPTON.

L'ozone atmosphérique, par CHARLES FABRY. Pp. 278. (Paris : Centre National de la Recherche Scientifique, 1950.) No price stated.

In the Preface to this posthumous production of Fabry's book, Cabannes writes : " A study of the concluding list of references shows that this survey of work on ozone ends in 1942. The book ought to be brought up to date some day even if, during the last seven years, progress has not been very rapid. But Fabry's approach will always catch one's interest. Geophysicists will continue to find here . . . the essentials of atmospheric optics. Techniques will improve . . . but the nature of the problems and principles behind the measurements do not change."

Here Cabannes has virtually summarized the scope of the book. The principles of atmospheric spectroscopy and the precautions which must be taken are very clearly set down, and in 1945, when Fabry died, even the *techniques* described were up to date so far as published work is in question.

However, the unavoidable delay in publication, which amounts effectively to eight years, has possibly dulled the interest of this book for the research worker now specializing in the study of atmospheric ozone. For him the book will certainly save a great deal of searching about in work published before 1942, since there must be very little significant work which Fabry does not summarize. Nevertheless, since measurements are based on techniques as well as guiding principles, the large technical advances during and since the war, particularly in light detectors, are rather a serious omission in a book only just now published.

On the other hand, for the postgraduate student beginning to study spectroscopy of any sort the book is a valuable guide to the rewards and limitations of his work. The important basic principles of spectrophotometry at low light levels, in the threatened presence of light scattered from more intense wavebands in the source, are most admirably pointed. It is very evident why Fabry was so successful a tutor. He has also mastered the rare art of being able to say " so-and-so has done such-and-such but . . . beware ! " almost wittily, without malice and always justifiably, and such remarks can be very valuable to students who have to read or carry out this sort of work for the first time with little supervision. They may be confident that, here, these observations are inspired by love of the subject and not by departmental pride.

Misprints are refreshingly few. Perhaps the most troublesome are : the lettering in figures 15 and 16 does not correspond to that in the text; p. 60, line 24 should read " au moins dans un certain domaine de valeurs de '*x*' "; p. 112, line 30 should read " $w=20\times10^{-9}$ " and p. 114, line 14, " $1,400\times10^{-9}$ "; p. 155, lines 19–20, the focal plane " P " of the text is not identified in figure 40; p. 156, lines 3 and 5, for " S_3 " read " S_5 "; p. 207, line 8, for " P " read " $P^{\frac{1}{2}}$ "; p. 210, bottom line, and p. 211, top line, for " figure 58 " read " figure 59 " and vice versa.

The promised reproduction of Ny Tsi-Zé and Choong Shin-Piaws' table of the absorption coefficients of ozone between 2136 and 3525 A. does not appear, and the footnote to page 169 is wrongly cross-referenced. Either reference 104, or perhaps Bemporade's values of air mass, are intended; the latter do not appear in the list of references.

There are a very few statements which might be misleading, e.g. on p. 158 it is not true that one can generally neglect $(\beta-\beta')/(\alpha-\alpha')$ even in the small wavelength range considered. Neither can the differences between μ, the light path length through the ozone layer, and m, through the whole atmosphere, be neglected at the large solar zenith distances at which measurements of atmospheric ozone can now be made with accuracy. Once again this is largely due to technical improvements since the war, in this case the successful production of multiplier photocells, and it is when these improvements in technique are not fully covered that some statements of general principle have no longer the correct emphasis. For the same reason measurements now possible at low sun make secondary and even higher order scattering important factors in interpreting the results. Yet the treatment of molecular and small particle scattering is really quite sketchy. Nor is it emphasized how nowadays photoelectric methods almost completely replace photography for the accurate determination of light intensity ratios.

The meteorology is possibly a little too confidently presented. For example, advection of polar and tropical air is probably only half the story of the causes of day-to-day increases and decreases in the ozone content of the atmosphere. Further, the enthusiastic researches of Vassy into correlating ozone changes with its effective temperature (measured spectroscopically) and applying the results to the theories of photochemical equilibrium of ozone and its variation with season and latitude are probably best taken very qualitatively.

However, the book is good to read and covers the whole field of ozone research, from determinations of total amount and vertical distribution, through theories of the formation and the causes of temporal changes of ozone, its influence upon radiative processes in the atmosphere, to finally even the biological effects of the striking opacity of ozone at the shortwave end of the solar spectrum. Mme Vassy has done very good work in editing the manuscript, for Fabry died before it was ready for publication.

Those interested in work done since 1942 should supplement the book by such publications as Dobson's Bakerian Lecture to the Royal Society (1946) and the appropriate sections of the forthcoming American compendium of meteorology.

The book is attractively produced but perpetuates the national tradition of not providing an index. It will be a valuable addition to any library and, if this review may be thought not sufficiently appreciative, then I take refuge in Fabry's own words (p. 113): " Il y a tout lieu de penser que l'auteur " (R.H.K.) " n'a pas su exprimer, dans sa langue, ce qu'il voulait dire " !

R. H. KAY.

Theory of the Interior Ballistics of Guns, by J. CORNER. Pp. xiii + 443. (New York: Wiley and Sons; London: Chapman and Hall, 1950.) 64*s*.

It is no coincidence that the lack of any book in English, of other than historical interest, dealing with internal ballistics, has been removed by the almost simultaneous publication of two—the volume now being considered, and the official compilation *Internal Ballistics* (H.M. Stationery Office), to which Corner was a contributor. Both books spring from the increased activity in internal ballistic research which was stimulated by World War II, and from the more enlightened policy which has since prevailed in permitting appreciable releases of the results of defence research. Although they cover the same general field, the emphasis and the method of approach differ widely, and there is plenty of room for both.

Dr. Corner first made contact with internal ballistics when he was with Sir John Lennard-Jones at the Cambridge University Mathematical Laboratory, where he approached the subject from the two different directions of the numerical solution of systems of internal ballistic trajectories on the differential analyser, and the mathematical chemistry of the burning of propellants. Subsequently, in the Armament Research Establishment of the Ministry of Supply, he himself made many contributions to the theory of the subject, and had full opportunities for discussions with all the other contributors in the same field. He is thus better qualified than most to write this book.

It was Hardy who said that not even Littlewood could make ballistics respectable. One wonders whether Corner has even tried to do so, but he has certainly succeeded in making the subject interesting. Whereas until comparatively recent times the theory of internal ballistics amounted to little more than a series of mathematical exercises based on a few items of faith which were only rough approximations to the phenomena depicted, the development of new experimental techniques has opened up a far wider view of the subject. As the author points out (and in so doing provides a much needed definition), "interior ballistics is an applied science; that is, it uses concepts evolved in pure science, instead of fashioning its own"; and he proceeds to show the contributions to the subject from modern arithmetic, hydrodynamics, thermodynamics, and from physical and theoretical chemistry.

The systems of internal ballistic calculation which were in general use until a few years ago were chiefly useful because of their simplicity and the ease with which they could be applied; but they were little more than semi-empirical methods for interpolating in the large mass of firing results which had been accumulated with orthodox guns and propellants; they were less successful when the higher performances required from guns called for extrapolation, or when applied to new types of propellants, or to the new and unorthodox weapons which bore some resemblance to guns. All the basic formulae of the old methods are examined and revised in the light of modern knowledge, more modern methods of ballistic calculation, which can take account of some, if not all, of these revised formulae, are described in detail, and the results are compared both with those obtained by the older method and with experimental data, a process which throws light on the adjusting constants found necessary with all systems. One feature which is not amenable to analytical treatment is the mechanical resistance to the motion of the projectile in the gun; this, in common with other complications, can, however, be dealt with arithmetically in special cases by such machines as a differential analyser, and a procedure for dealing with the results by means of weighting factors is borrowed from the allied science of external ballistics.

The burning of propellants both in closed vessels and in guns is examined in considerable detail, and the point is made that the value of a true theory of burning lies not so much in the ability that it gives to predict rates of burning (these can be obtained either empirically from a knowledge of the propellant composition, or by routine experiments), but that it throws light on phenomena associated with the burning, such as 'erosive burning', or ignition. Surface theories of burning are inadequate for this purpose, and are not in accord with modern views of chemical reactions; by treating burning as a vapour phase reaction, and by allowing for the effects of diffusion from the flame zone, in the light of these views, a satisfactory qualitative and quantitative agreement with observed phenomena is achieved. Modern ideas from theoretical chemistry are also applied to the classical thermochemistry of propellants to explain the difference between the ‘ force ’ of the propellant in a closed vessel and in a gun, the value of the ‘co-volume’ and the variations in the ratios of the specific heats of the gases.

The hydrodynamics of the propellant gases are considered in a detailed study of the pressure gradient along the barrel, which is becoming of increasing importance as muzzle velocities tend to rise. In addition the Rateau–Hugoniot work on the emptying of the gases from the gun is re-examined, extended, and applied to such problems as the residual gas pressure in the gun chamber after shot ejection, and the design of muzzle brakes, and methods are described for comparing the efficiencies of muzzle brakes. Other hydrodynamical problems which are studied include the motion of the air ahead of the projectile, and the motion of the unburnt propellant. It is deduced that the solid propellant remains in or near the gun chamber until burning is nearly complete, but the last remaining portions may move forward with a high acceleration.

One of Corner's more important contributions to the subject is his ‘ theory of leaking guns ’. By considering the flow of gas through a nozzle he arrives at a theory for the cases where the gun bore and chamber are not completely sealed by the breech mechanism and the projectile, and by the introduction of one additional parameter to deal with the leakage he derives a series of approximate equations for routine working which are little more complicated to use than the simplest set of semi-empirical equations for orthodox guns. This theory and these equations are applied to the study of such problems as recoilless guns, smooth bore mortars, or orthodox guns which are so worn that gas can escape past the projectile.

The last chapter of the book describes Hicks and Thornhill's work on the transfer of heat from the gases to the barrel wall of the gun.

In a stimulating section the author suggests the lines on which further theoretical research is needed : the further applications of present theories to new forms of weapons, or to cases where there is substantial resistance to the motion of the shot, chemical problems set by the ignition and burning of propellants, the study of plasticized polymers, and the hydrodynamics of the gun.

The book, and its title, have been painstakingly translated into American by the publishers. G. H. H.

CONTENTS FOR SECTION A

PAGE
Mr. P. N. Butcher. The Absorption of Light by Alkali Metals . . . 765
Miss M. E. Pillow. Transition Probabilities in Band-Systems of Diatomic Molecules : A Modified ‘Distortion’ Process for the Wave Functions . . . 772
Dr. S. Levine. The Free Energy of the Double Layer of a Colloidal Particle . . . 781
Dr. D. K. Butt and Mr. W. D. Brodie. The Conversion and Auger Electrons of Radium D 791
Dr. J. Hughes. The Angular Distribution of Neutrons from the (d, n) Reaction in Light Elements . . . 797
Mr. R. Middleton and Mr. C. T. Tai. Deuteron Bombardment of Neon . . . 801
Mr. H. Messel. Further Results on the Fluctuation Problem in Electron–Photon Cascade Shower Theory and the Probability Distribution Function . . . 807
Mr. A. B. Lidiard. On the Theory of Free Electron Ferromagnetism . . . 814
Dr. K. R. Atkins and Mr. C. E. Chase. The Velocity of First Sound in Liquid Helium . 826
Dr. K. R. Atkins. The Flow of Liquid Helium II through Wide Capillaries . . . 833
Letters to the Editor :
  Mr. P. Gay and Dr. R. W. K. Honeycombe. X-Ray Asterisms from Deformed Crystals 844
  Dr. W. T. Davies and Dr. M. A. Grace. On the Identity of β-Rays with Electrons . 846
  Mr. D. B. James and Mr. P. B. Treacy. Measurement of Fast Neutron Velocities by Delayed Coincidence . . . 847
  Mr. A. J. Salmon. An Investigation of the Reaction ^{9}Be(d, p) . . . 848
  Mr. J. J. Wilkins and Mr. F. K. Goward. Note on Experimental Determination of the Half-Life of ^{8}Be . . . 849
  Mr. Suraj N. Gupta. On the Supplementary Condition in Quantum Electrodynamics 850
  Mr. S. G. Krishnamurty. The Band Systems of Thorium Oxide and Hafnium Oxide 852
Reviews of Books . . . 852
Contents for Section B . . . 857
Abstracts for Section B . . . 858

ABSTRACTS FOR SECTION A

The Absorption of Light by Alkali Metals, by P. N. Butcher.

ABSTRACT. The theory of the absorption of light by photo-excitation of the valency electrons in cubic metals is developed. The matrix elements vanish when calculated with respect to the Wigner–Seitz approximate wave functions for the occupied states in the first band, and free electron wave functions for the states in the higher bands. Calculations of the contribution to the absorbing powers of the alkali metals due to photo-excitation are made, using the approximation of nearly free electrons to obtain all the unperturbed wave functions. The calculated values for sodium, potassium and rubidium agree quite well with those observed if the first non-vanishing Fourier coefficient of the self-consistent potential field has the values 0·323, 0·305 and 0·350 ev. respectively.

Transition Probabilities in Band-Systems of Diatomic Molecules: A Modified 'Distortion' Process for the Wave Functions, by M. E. Pillow.

ABSTRACT. A method previously used for 'distorting' the wave functions of a simple harmonic oscillator to fit the potential curve for a particular electronic state has been modified, giving wave functions for the vibrational states which are close approximations to those of Morse, with comparatively little labour. The transition probabilities obtained from these functions are compared with experimental values for the first negative bands of N_2^+, the Swan bands of C_2, the CN violet bands, and the α-system of BO.

The Free Energy of the Double Layer of a Colloidal Particle, by S. Levine.

ABSTRACT. Some theorems concerning the free energy of the electric double layer of a colloidal particle in a dilute sol are developed by making use of the Poisson–Boltzmann equation. It is shown that this free energy can be derived by employing the fictitious process of charging the excess ions in the double layer and that the surface density of ions may be an arbitrary function of the degree of charging. By integrating with respect to the charging parameter at constant surface density of ions we obtain an alternative form for the free energy which is expressed in terms of the electrostatic energy of the double layer and the excess osmotic pressure in the diffuse layer. The latter expression for the energy can also be derived by attributing to each volume element in the diffuse layer an energy term due to the thermal motion of the ions and an (electrostatic) field energy term.

The Conversion and Auger Electrons of Radium D, by D. K. Butt and W. D. Brodie.

ABSTRACT. The conversion and Auger electrons of Radium D have been studied in a lens β-ray spectrometer with a post-focusing electron accelerator. The number of conversion electrons due to the 46·7 kev. γ-ray is found to be $59{\cdot}2 \pm 5$ per hundred disintegrations but reasons are given which suggest that this value be regarded as a lower limit. The Auger electrons have been compared with those resulting from the ThB–ThC transition.

The Angular Distribution of Neutrons from the (d, n) Reaction in Light Elements, by J. Hughes.

ABSTRACT. A proton recoil chamber with a thick polythene layer was used to determine the angular distribution of the neutrons from thick targets of various light elements bombarded with 8 Mev. deuterons. In all cases a pronounced forward maximum in intensity was found, the half width of the distribution decreasing with increasing atomic number. The absolute intensity of neutrons in the forward direction decreases as the atomic number increases but the dependence on atomic number is not a smooth function. Most elements show the presence of multiple forward peaks. In a few elements the effect of reducing the incident deuteron energy was investigated.

Deuteron Bombardment of Neon, by R. Middleton and C. T. Tai.

ABSTRACT. Neon gas of normal isotopic constitution was bombarded by 7·8 Mev. deuterons from the Liverpool University cyclotron. Photographic plates were used as detectors, enabling the protons from the reactions ^{20}Ne(d, p) and ^{22}Ne(d, p) and alpha-particles from ^{20}Ne(d, α) and ^{22}Ne(d, α) to be studied, as well as a group of inelastically scattered deuterons.

Further Results on the Fluctuation Problem in Electron–Photon Cascade Shower Theory and the Probability Distribution Function, by H. Messel.

ABSTRACT. Results are presented both for the fluctuation and distribution of electrons in electron–photon cascade shower theory. The results were obtained in approximation A and are for the case of a single primary electron giving rise to the shower.

On the Theory of Free Electron Ferromagnetism, by A. B. Lidiard.

ABSTRACT. A calculation has been made of the ferromagnetic properties of a system of electrons using plane wave functions and Bloch exchange integrals. Our results show that below a critical temperature T_c the spontaneous magnetization decreases slowly with increasing temperature. At T_c the magnetization drops suddenly to zero as if a change of phase were taking place and there is therefore a latent heat of demagnetization at this temperature. Above T_c the model is paramagnetic. The reasons for the differences between the results and those obtained by experiment and on Stoner's theory of collective electron ferromagnetism are discussed, and the possibility that they may be due to the neglect of positional correlations among the electrons is considered.

The Velocity of First Sound in Liquid Helium, by K. R. Atkins and C. E. Chase.

ABSTRACT. The velocity u_1 of ordinary sound in liquid helium has been measured in the temperature range from 1·2° K. to 4·2° K. Its value extrapolated to 0° K. is compared with the velocity u_2 of second sound at 0·1° K. and a slight departure from the Landau relation $u_2=u_1/\sqrt{3}$ is indicated. The behaviour of u_1 in the neighbourhood of the λ-point has been studied in detail.

The Flow of Liquid Helium II through Wide Capillaries, by K. R. Atkins.

ABSTRACT. A study has been made of the isothermal, pressure-induced flow of liquid helium II through glass capillaries of bore between $2{\cdot}6\times10^{-3}$ and 4×10^{-2} cm. The critical velocity is found to be less than 1 cm. sec^{-1} except, possibly, in the case of the finest capillary. The theory that there is a force of mutual friction between the normal and superfluid components predicts the correct order of magnitude for the flow velocities, but there are discrepancies which suggest the existence of some other source of friction.

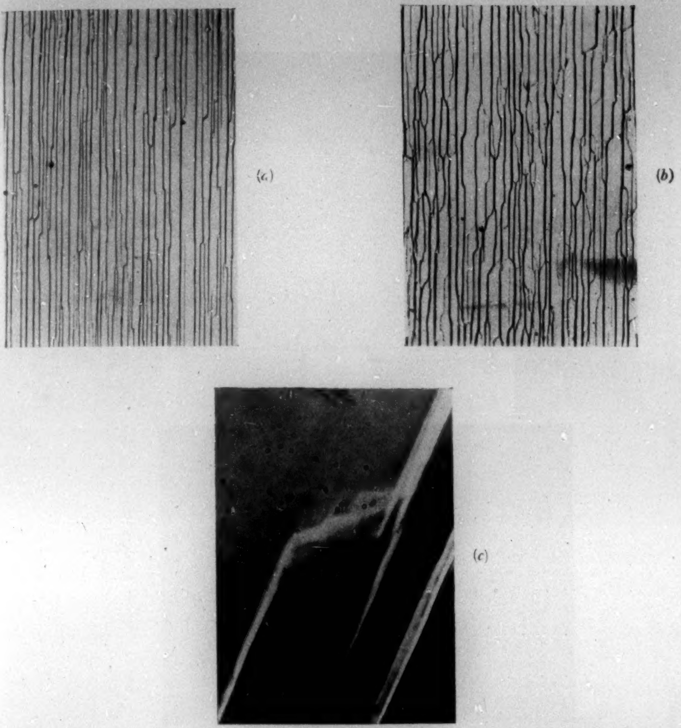


Figure 9. Cross slip on top surface of aluminium crystals. (*a*) Crystal extended 7% at room temperature. (×94) (*b*) Crystal extended 9% at 500° c. (×82). (*c*) Crystal extended 5%. (×1760)

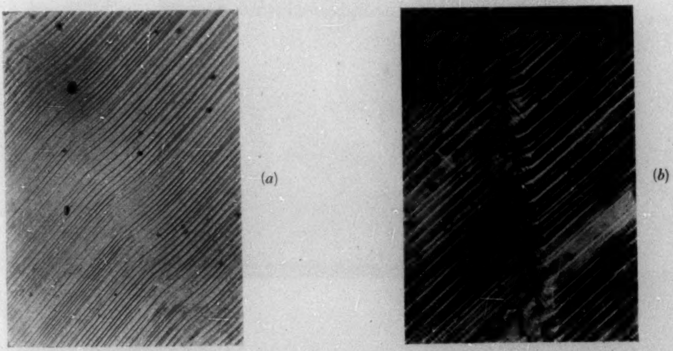


Figure 10. Slip on side surface. (*a*) A deformation band; crystal extended 20%. (×188) (*b*) A polygonized deformation band; crystal extended 31%, annealed one hour at 450° c. (×150)

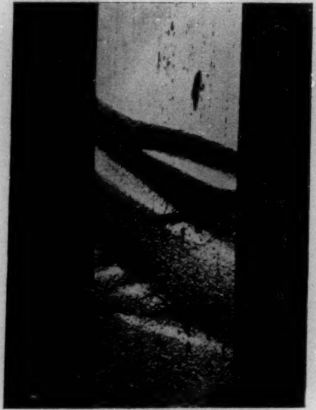


Figure 1. Complex Lüders bands. (×4)
(*Photo by courtesy of Mr. W. M. Lomer.*)

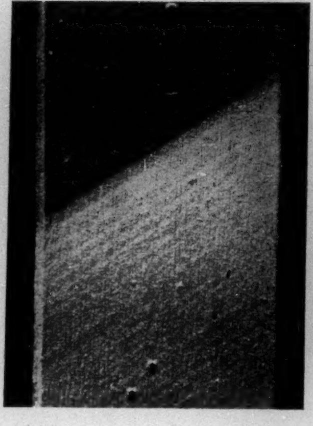


Figure 2. Simple Lüders band. (×3½)

Figure 3. Angular deviation at the junction of two bands. (×1½)

PLATE I.

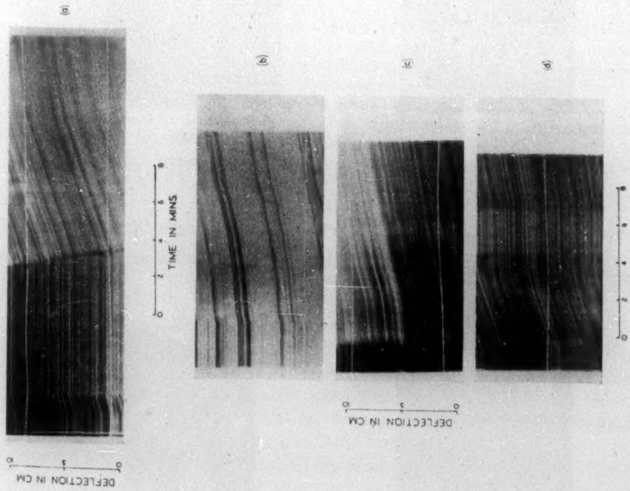


Figure 7. Optical traces from virgin material. (×⅓)

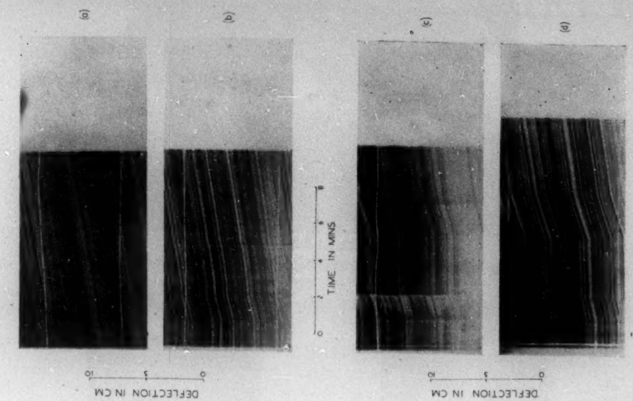


Figure 8. Optical traces from strain-aged material. (×⅓)

Plate II.

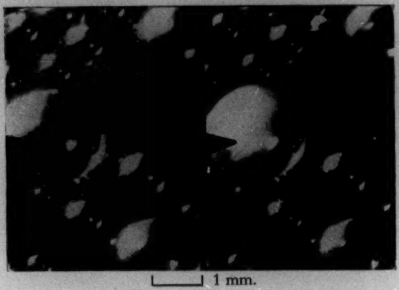
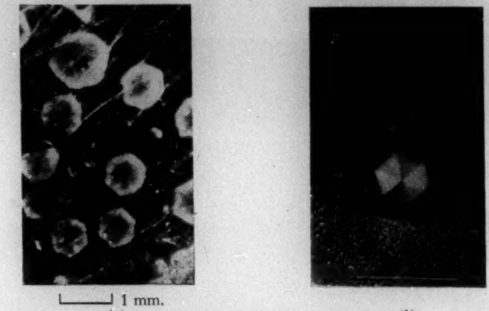
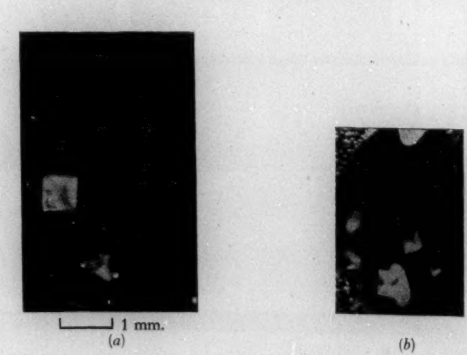


Figure 3. The freezing of a drop of water photographed at 64 frames per second. The drop goes white between two frames.



(a) (b)

Figure 6. (a) Hexagonal crystals in a water drop which has been contaminated with alcohol. (b) Photograph published by Rau.



(a) (b)

Figure 7. (a) Cubic crystals appearing at −72° c. in a water drop which has been contaminated with alcohol. (b) Photograph published by Rau.